DE GRUYTER
OPEN

STUDIES IN LOGIC, GRAMMAR
AND RHETORIC 40 (53) 2015

DOI: 10.1515/slgr-2015-0004

Paweł Gładziejewski
Institute of Philosophy and Sociology, Polish Academy of Sciences

EXPLAINING COGNITIVE PHENOMENA WITH INTERNAL REPRESENTATIONS: A MECHANISTIC PERSPECTIVE

Abstract. Despite the fact that the notion of internal representation has – at least according to some – a fundamental role to play in the sciences of the mind, not only has its explanatory utility been under attack for a while now, but it also remains unclear what criteria should an explanation of a given cognitive phenomenon meet to count as a (truly, genuinely, nontrivially, etc.) *representational* explanation in the first place. The aim of this article is to propose a solution to this latter problem. I will assume that representational explanations should be construed as a form of mechanistic explanations and proceed by proposing a general sketch of a functional architecture of a representational cognitive mechanism. According to the view on offer here, representational mechanisms are mechanisms that meet four conditions: the structural resemblance condition, the action-guidance condition, the decouplability condition, and the error-detection condition.

Keywords: mechanistic explanation, representationalism, antirepresentationalism, mental representation, s-representation, emulation theory, predictive coding.

1. Introduction

The existence and explanatory utility of mental representations remains a hotly debated topic at the intersection of philosophy and cognitive science. Are representations a useful explanatory tool in the sciences of the mind, or are they of purely historical value and ought to give way to new, perhaps more sophisticated tools? Interestingly, it seems that in recent years, after decades of representationalism's dominance, the tide has been steadily turning in favor of the antirepresentationalist position (for some recent attempts to formulate and defend antirepresentationalism, see Calvo, 2008; Chemero, 2009; Hutto & Myin, 2013).

Although the present paper is devoted to the problem of the explanatory role of mental representations in cognitive science, my aim will *not* be to argue for either representationalism or antirepresentationalism. I think that

ISBN 978-83-7431-440-4 ISSN 0860-150X

it remains largely unrecognized, or underappreciated, that there is not one, but *two* important theoretical problems regarding the explanatory status of representations in cognitive science. The first, rather obvious problem is that of whether the notion of internal representation is actually useful in our attempts to explain cognitive phenomena. Let us call it the "first-level problem". However, the second problem – the one whose importance, I think, is not appreciated enough in the literature – pertains not to whether representations should figure in (true) explanations of cognition, but to the question of *how we should even construe the very nature of representational explanations.* Let us call this latter issue the "second-order problem". The second-order problem can be expressed in questions such as these: What exactly makes a given cognitive-scientific explanation a representational explanation? What does it mean for a representation to play an explanatory role? What distinguishes a situation in which a given cognitive phenomenon is actually explained by internal representations from a situation when the representational terminology is a mere non-explanatory gloss that adds nothing to what is at heart a non-representational explanation (see Chemero, 2009; Ramsey, 2007)?

My aim in this paper will be to propose a solution to the second-order problem of representation. There are (at least) two reasons to think that this is a worthwhile pursuit. First, at this point we simply lack a clear, well-justified account of what representational explanation amounts to in cognitive science (Haselager, de Groot & van Rappard, 2003; see also Ramsey, 2007). This is not only unfortunate in itself, but also stands in the way of our attempts to resolve the first-level problem: it seems impossible to resolve the representationalism/antirepresentationalism debate conclusively unless we have at our disposal a principled, universally agreed upon understating of representational explanation. Second, we cannot simply bypass the second-order problem by saying that representational explanations are whatever cognitive scientists routinely treat as representational explanations in their actual scientific practice. As William Ramsey (2007) forcefully argues, it is oftentimes far from clear whether an internal structure postulated as a representation by researchers actually plays – in some illuminating, explanatorily valuable way – the role of *representation.* In fact, the same author argues that the notions of representation most commonly used in current cognitive science do *not* pick out cognitive structures that are worthy of the representational status.

In my attempt to solve to the second-order problem of representation, I will draw from recent developments in the philosophy of science. More specifically, I will develop my proposal against the background of the mechanistic model of scientific explanation, as applied to cognitive-scientific explanation (see Bechtel, 2008; Piccinini & Craver, 2011). The plan is to propose a set of conditions, such that if a mechanistic explanation of some cognitive phenomenon meets those conditions, it should count as a (genuinely) *representational* mechanistic explanation.

The article will have the following structure. In the second section, I will briefly present the mechanistic account of explanation and draw from it some general conclusions with respect to the problem of representational explanation. In the third section, I will present a solution to my problem of interest by proposing a general functional sketch of a representational mechanism. In the third section, I will address two possible worries regarding my proposal. I will close the paper by drawing general conclusions from the discussion presented here.

2. Representational explanations as mechanistic explanations

As mentioned above, before tackling the topic of representational explanation, I want to make a short excursion into the philosophy of science to get a clearer idea about the nature of explanation in cognitive science as such. Throughout this paper, I will assume that (1) the basic or dominant[1] type of explanation in cognitive science is *mechanistic* explanation, and that (2) representational explanations in cognitive science should be construed as a species of mechanistic explanation. Let me start with a brief presentation of the mechanistic model of scientific explanation.

Following the accounts of mechanistic explanation that have gained considerable recognition and popularity in recent philosophy of science, I take the practice of explaining phenomena by their mechanisms to consist in the structural and functional decomposition of a complex system to show how its causal architecture at the lower level of organization enables the system as a whole to exhibit a certain capacity (for influential presentations of the modern version of mechanism, see Bechtel, 2008; Bechtel & Abrahamsen, 2005; Craver, 2007; Glennan, 2002; Machamer, Darden & Craver, 2001). More precisely, according to mechanism, explaining a phenomenon of interest – understood as a capacity of some complex system – consists in discovering its mechanism, which in turn consists in discovering component parts of the system in question, discovering the activities or functions[2] that these component parts perform, and then showing how the organized functioning of those component parts brings about the phenomenon. From such a point of view, explanation has a (reverse-)engineering flavor: instead of answering why-questions ("Why does X occur?"), mechanistic explanations should be seen as providing answers for how-questions ("How does a system S, exhibiting capacity Y, work?").

Although mechanism's most obvious applications lie within the philosophy of biology, it seems that the explanatory practice of cognitive scientists can also be captured by the mechanistic model (see Bechtel, 2005; Piccinini & Craver, 2011). What cognitive science aims to do, after all, is to explain cognitive capacities – like episodic memory, visual perception, concept acquisition, reasoning, sensorimotor integration, imagery, or mindreading – exhibited by complex, human or non-human cognitive systems. In order to explain those capacities, cognitive scientists look for their underlying mechanisms: sets of organized, functioning parts of a cognitive system – presumably located inside its head (Bechtel, 2005) – that jointly give rise to those phenomena. As John McDowell (1994) rightly points out, the job of cognitive science is to discover lower-level, mechanistic *enabling conditions* of system-level cognitive phenomena.

I propose that representational explanation in cognitive science is also a sort of mechanistic explanation. That is, cognitive scientists who postulate internal representations are involved in the project of searching for mechanisms that underlie cognitive phenomena. Since representations are usually understood as internal states or structures somehow located inside the cognitive system – that is, they are understood as component parts of the cognitive system – I think it is natural to treat representational explanations as mechanistic. If this is so, then contrary to what is more or less explicitly assumed in the philosophical literature, *scientific* representational explanations are *not* "naturalized" versions of folk-psychological explanations, with the latter construed as either some sort of covering-law (à la Fodor, see e.g. Fodor, 1994) or causal-etiological explanations (à la Dretske, see Drestke, 1988; for the distinction between causal-etiological and mechanistic or "constitutive" explanations, see Craver, 2007). This means that representations figure as explanantia in mechanistic explanations that answer how-questions about cognitive capacities, and *not* in explanations that answer why-questions about particular behaviors or actions. Representations in this sense do not play the explanatory role traditionally attributed to propositional attitudes.

But what could representational mechanistic explanations possibly be? What sort of explanatory role do representations play from the perspective of mechanism? Since mechanisms are comprised of organized, functioning components, it seems natural to equate representations with (postulated) functionally specified components of some (postulated) cognitive mechanism. Consider this simple analogy (see also Ramsey, 2007): When a scientist postulates that blood is circulated through the body by a pump, the natural way to interpret this claim is to think that the mechanism of blood circulation is comprised of (among others) a component part that functions as a pump. The same way, if a cognitive scientist postulates internal representations to explain some phenomenon, the natural way to interpret this mechanistically is to think that the mechanism responsible for that phenomenon is comprised of (among others) a component part that functions as a representation. Let us simply call a mechanism of this sort a "representational" mechanism. To explain a phenomenon representationally, then, is to explain it by a representational mechanism. Thus, looking at the second-order problem of representation through the lens of mechanism leads us to two basic general conclusions:

(1) A mechanistic explanation M of a cognitive capacity C is representational iff M explains C by a representational mechanism.

(2) A mechanism M is representational iff M has at least one component part whose function within the mechanism consists in representing something.

This is where things become more complicated. It is easy to say that representations are component parts of mechanisms that play the functional role of a representation. But it is much harder to answer the question of *what it means to function as a representation within a mechanism.* When are we justified in attributing the role of a representation to a component of a neural or computational mechanism? What *exactly* does a component have to *do* within a mechanism in order to be justifiably categorized as a representation? To have a theory of mechanistic representational explanation, we need a principled answer to those sorts of questions.

To get a clearer idea about the issue at hand, let me briefly discuss Ramsey's immensely useful book *Representation Reconsidered* (2007). In it, the author argues that structures routinely posited as representations by cognitive scientists should be confronted with what he calls the "job description challenge" (henceforth JDC). Meeting the JDC requires showing, in detail, how or in what sense a structure that features as a representation in a given cognitive model or theory in fact does serve a truly representational role inside the cognitive system; for example, in what sense does this structure serve as a *stand-in* for some external state of affairs, analogously to the way things we pretheoretically recognize as representations – like maps or fuel gauges – stand-in for other things. Only if the JDC is met can we be justified in saying that a given structure actually plays the functional role of a genuine representation (as opposed to being simply called "representation" when in fact playing a non-representational role). And only then can we say that the explanation that features this structure as an explanans, or part of an explanans, makes use of a notion of representation that plays a genuine, non-trivial explanatory role. If the JDC is not met, then we are dealing, according to Ramsey, with a non-representational explanation misleadingly dressed in representational talk.

Ramsey's (2007) project is to (1) delineate different notions of representation routinely used by cognitive scientists and (2) evaluate the explanatory value of those notions by verifying whether the cognitive structures they designate meet the JDC. Although it is impossible to discuss Ramsey's analysis in detail here, suffice it to say that if he is right then some of the most popular notions of representation in cognitive science do *not* pick out internal states or structures that play representational functions in any clear, intuitively recognizable way (and therefore do not meet the JDC). If this is right, most "representational" explanations in cognitive science are representational in name only. For example, Ramsey argues at length that the often used "receptor" notion of representation – which construes representations as detectors that reliably co-vary with some worldly state of affairs – turns out to equate serving as a representation with being a reliable causal mediator. Since representing something is not the same as functioning as a causal mediator, the receptor notion, according to Ramsey, picks out structures that do not meet the JDC, and thus is explanatorily vacuous.

I agree with Ramsey that the JDC is a useful tool for dealing with the question of representational function. And since the question of representational function lies at the heart of the problem of representational mechanisms, the JDC may also play a key role in my project. The moral to draw from Ramsey's work is that one cannot formulate a theory of representational mechanisms by either simply proclaiming a given class of mechanisms as representational, or by looking at what sorts of mechanisms are actually described as representational in cognitive-scientific practice. A mechanism needs to *earn* its representational status by virtue of having a functional and structural architecture such that we can show that (at least) one of its component parts meets the JDC and thus functions as a representation within this mechanism. My objective in the remaining part of the paper will therefore be to develop a job description for a component part of a possible mechanism, such that if that component meets that job description, then it can be attributed – in a justified and explanatorily beneficial manner – the role of representation within a larger mechanism, thus earning this larger mechanism the status of a truly *representational* mechanism.

3. Representational mechanisms

3.1. Some preliminary remarks

Before I move on to set out my proposal, let me start with some introductory clarifications. First, it is important to note at the outset that what I will offer here is merely a general, incomplete, and highly idealized *sketch* of a *possible* mechanism (see Craver, 2007; Piccinini & Craver, 2011). This sketch will be formulated in purely functional terms and will abstract away from structural details, i.e. from whether there are actual mechanisms in real cognitive systems whose component parts correspond to the functional organization I am proposing (although, see section 4). This is not a weakness of the present proposal. The idea, after all, is to show what representational explanation amounts to, and not to provide representational explanations for specific phenomena. What I claim, then, is that my functional sketch – although it is not a full-blown explanation by itself – captures the general nature of representational mechanistic explanations. Any *actual* mechanistic explanation of some cognitive phenomenon should count as representational if the mechanism it postulates has component parts that correspond to my functional sketch; or, in other words, if it can be shown that the explanation in question fills in my functional sketch with structural details (for an illuminating discussion of the explanatory role of purely functional models in cognitive science from the point of view of mechanism, see Piccinini & Craver, 2011).

Second, my aim here is not to put forth a *definition* of representation. Although I will propose a set of conditions for being a representational mechanism, these should not be understood as individually necessary and jointly sufficient conditions. I agree with Ramsey (2007) that attempts to define representation are probably futile and that the concept of representation is probably prototypical. So the idea is, rather, to propose a set of conditions that are jointly sufficient for being a representational mechanism because meeting them guarantees an appropriate level of similarity to prototypical cases of representation-use (see section 4).

Third, what I aim to do here is not to invent some revolutionarily new way of thinking about representation, but rather to integrate – within the mechanistic framework – some ideas that are already present in the literature. More specifically, the goal is to put together two seemingly incompatible ways of thinking about the nature of representation: one which construes representations in terms of a relation ("correspondence") between the representation itself and what is represented, and the other which construes representation in terms of (inter)action-guidance that representation provides for its user (see Anderson & Rosenberg, 2008; Bickhard, 1999, 2004). The idea is that integrating those two approaches will hopefully provide us with a notion of representation robust enough to meet the JDC (for a more detailed description of theoretical motivations behind pursuing this kind of approach, see Gładziejewski, 2015; for a proposal that is in many aspects close to mine, see Miłkowski, 2013).

Fourth, by endorsing mechanism, I will not follow Anthony Chemero's (2009) advice to distinguish and keep separate the *epistemological* problem of representation (pertaining to whether the best explanations of cognitive phenomena are representational) and the *metaphysical* problem of representation (pertaining to whether representations actually exist and are responsible for cognitive phenomena). If Chemero were right, then it could well turn out that either (1) representational explanations are the best available for a cognitive scientist, but representations do not really exist and their explanatory value is purely instrumental, or (2) representations do exist and are responsible for cognitive phenomena, but non-representational explanations are preferable to representational ones. Both these options are incompatible with the mechanist view of explanation. Regardless of *what* exactly counts as an explanation – the worldly mechanism (as proponents of the ontic interpetation of mechanism claim, see e.g. Craver, 2007) or its scientific representation (as the proponents of the epistemic interpretation hold, see e.g. Bechtel, 2008) – mechanists universally claim that the *accuracy* of any mechanistic explanation depends on objective facts about the world, namely on the nature of the mechanism that actually underlies a phenomenon that is being explained (Bechtel, 2008). The epistemological and metaphysical factors cannot be neatly separated because explanations need to capture the actual mechanistic structure of the world; having predictive power or even being counterfactual-supporting is not enough to be a good explanation. This also applies to representational mechanistic explanations. The accuracy of any representational explanation rests on whether the explanandum phenomenon is really underpinned by a representational mechanism. For this reason, representations cannot feature in accurate explanations of phenomena unless they are responsible for those phenomena; and *vice versa*: representations cannot be responsible for phenomena without featuring in accurate explanations of those phenomena. This is why the conception I am about to put forward is *both* a theory of what representational *explanations* are and what representations (representational mechanisms) *themselves* are.

With these preliminary issues out of the way, let me get to the point. I propose that a representational mechanism is a mechanism that meets four conditions regarding its functional organization: the structural similarity condition, the action-guidance condition, the decouplability condition, and the error-detection condition. In sub-sections 3.2–3.5, I discuss each of these conditions in turn.

3.2. Structural resemblance condition

As mentioned above, from the point of view of mechanism, representations – if they exist at all – are functioning (working) parts of cognitive mechanisms. Let us call a component part of a mechanism that is functionally involved in representing something a "representational vehicle". The question, then, is what sorts of things should a component of this sort *do* in order to earn the status of a representational vehicle.

My first constraint on a representational vehicle is connected to the relation that holds between the vehicle itself and what it represents, or its representational object. According to the view I am advocating, this relation is *structural resemblance* (see Bartels, 2006; Cummins, 1989; O'Brien & Opie, 2004; Ramsey, 2007; Swoyer, 1991). That is, the vehicle should resemble that which it represents, not in its physical properties (e.g. color), but rather in its *structural organization* (see the distinction between first- and second-order resemblance in O'Brien & Opie, 2004). In other words, the pattern of relations among the constituents of the vehicle should mirror the way the represented object is relationally organized.[3] Gerard O'Brien and Jon Opie (2004, p. 11) characterize structural resemblance more technically:

> Suppose $SV = (V, \mathfrak{R}V)$ is a system comprising a set V of objects, and a set $\mathfrak{R}V$ of relations defined on the members of V.... We will say that there is a second-order [structural – PG] resemblance between two systems $SV = (V, \mathfrak{R}V)$ and $SO = (O, \mathfrak{R}O)$ if, for at least some objects in V and some relations in $\mathfrak{R}V$, there is a one-to-one mapping from V to O and a one-to-one mapping from $\mathfrak{R}V$ to $\mathfrak{R}O$ such that when a relation in $\mathfrak{R}V$ holds of objects in V, the corresponding relation in $\mathfrak{R}O$ holds of the corresponding objects in O.

Thus, representational mechanisms as I understand them employ so called *structural representations*, or s-representations (Cummins, 1989). Notice that the definition above does not require the relation between the vehicle and what it represents to be a full-blown isomorphism: one-to-one mapping should hold for *at least some* constituents of the vehicle and *at least some* of the relations between them. The s-representation in this sense does not need to be a complete structural copy of the represented object. Some weaker version of structural similarity – like homomorphism or "embedded isomorphism" (see Swoyer, 1991) – will suffice. This way of thinking is perfectly compatible with the idea that internal representations might – and probably will – turn out to be sketchy or idealized, capturing only *some* structural aspects of the represented object.

One important problem might be raised in the context of the structural resemblance condition. Although structural similarity is a relational property of the representational vehicle, it is not a functional property. Given that what I am looking for is some theory of the *function* that representations perform in cognitive mechanisms, the appeal to a non-functional property might seem unhelpful.

To answer this worry, it is useful to distinguish (1) a relation that simply may or may not hold between the representational vehicle and the representational object (a relational property the vehicle may or may not have), from (2) a relation *which is relevant to whether the vehicle fulfills its function inside a larger mechanism* (a relational property such that the vehicle functionally depends on having this property). By appealing to structural similarity, I mean that it plays this latter sort of role. Thus, a representational component of a mechanism is not simply a component that is structurally similar to something in the world. Rather, a component part of a mechanism gets to be a representational vehicle if its success at playing its function inside the mechanism is systematically and non-accidentally – that is, in virtue of the causal-functional architecture of the mechanism – dependent on whether there is a structural resemblance between this component and some worldly state of affairs (the representational object). In simpler terms, the vehicle "needs" to structurally resemble the representational object in order to *function* properly in the mechanism. What I mean by this will become clearer in the next sub-section. But the point to make here is that although structural resemblance itself is not a functional property, it turns out to be a property that is relevant to the functioning of the vehicle within a larger mechanism; it is a relation that is *exploitable* for the mechanism (see also Shea, 2007).[4]

3.3. Action-guidance condition

The question now is, of course, *what sort of function* should a component part that constitutes a representational vehicle play, such that fulfilling this function is systematically and non-accidentally dependent on structural resemblance? To put it more simply, what is the function to which the resemblance is relevant? This is where my second condition for being a representational mechanism comes into play.

Notice first of all that mechanism is a natural ally of the idea that representations should be construed not only as representations *of* some representational object, but also as representations *for* some sort of representation-user. Within the mechanistic framework, to be a representation is to be *used* as representation in a mechanism, or to be, in some sense, a representation *for* a larger mechanism. So what we are dealing with is not a two-party relation between the representational vehicle and the represented object, but a three-party relation between the vehicle, the object, and some sort of representation user.

I propose that the function of a representation within a representational mechanism is *action guidance* (see Anderson & Rosenberg, 2008; Bickhard, 1999, 2004; Miłkowski, 2013). Representations in this sense are in the business of driving the actions of larger cognitive mechanisms – and so, indirectly, of cognitive systems as wholes – with respect to external (represented) states of affairs. I borrow this general concept from Mark Bickhard's (1999, 2004) interactivist theory of representation. Let me illustrate this idea with an extremely simple example of motor control. Imagine a fictional organism which is equipped with a motor control mechanism capable of generating two types of movement, R1 and R2. One important point to make is that the activity of M should be analyzed within the context of the organism's external environment, since its *success* in controlling the organism's actions depends on the conditions of this environment. More specifically, imagine that there are separate types of environmental conditions that correspond to the two aforementioned types of movement. Let us say that R1 contributes to M's success under environmental conditions R1', and R2 does the same under (non-overlapping) environmental conditions R2'.[5] Now, the point is that representations come into play when control structures such as M – structures whose success depends on external conditions – do not have any sort of direct causal contact or coupling with those conditions, and thus their activity cannot be guided by the world itself, so to speak.[6] Representations are a mechanistic way of dealing with this problem. They mediate between the external states of affairs and (parts of) the internal, mechanistic architecture of a cognitive system. They stand-in for the world and adapt the activity of internal mechanisms to their worldly conditions of success.

Let me try to unpack this general idea further using more explicitly mechanistic terms. Although the vehicles of internal s-representations I am postulating can be attributed the role of providing guidance for action, they do not perform this function by interacting with a cognitive system as a whole. That is, they do not perform their role the same way that external s-representations – like maps or diagrams – perform it, viz. by being interpreted and employed as representations by (human) cognitive systems as such. Rather, representations construed as component parts of mechanisms play this role indirectly, in virtue of their causal/functional relationships with *other component parts*.[7] More specifically, and borrowing from Ruth Millikan (1984), I propose that internal representations guide action in the sense that inside the mechanism there is another component part that is a representation *consumer*.[8] A representation consumer in the sense I am advocating is a component of a (representational) mechanism that (1) has a function (performs an activity) in the mechanism, and being successful at fulfilling this function depends on states of affairs outside the mechanism, so that the consumer needs to work in a way that is adapted to those states; (2) is not causally coupled with those states of affairs, but is causally coupled with the representational vehicle; (3) is functionally dependent on the representational vehicle, in the sense that it functions properly under the condition that the vehicle actually structurally resembles the external states of affairs. In other words, the consumer is a component part of a mechanism that benefits from a situation where the representational vehicle resembles what is represented. The diagram below shows a general outline of this Peirce-style triadic relationship.

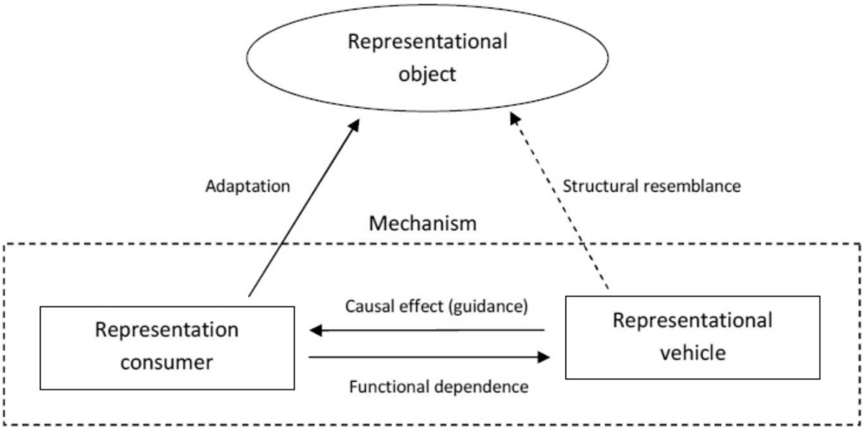

Figure 1. A schematic diagram showing the relationships between the representational vehicle, the representation consumer, and the representational object (see main text for details)

For an illustration, consider the toy example discussed by Ramsey (2007), who in turn draws it from Robert Cummins (1996). Imagine a self-driving car which faces the challenge of navigating its way through an S-shaped track, and which succeeds at this in the following way:

> One way we might do this [i.e. automatize the way the car works – PG] ... would be to convert the S-curve of the map [of the track – PG] into an S-shaped groove into which a rudder would fit. The rudder could then move along the groove as the vehicle moves forward, and the direction of the steering wheel and, thus, the vehicle's front wheels could be made to correspond to the direction of the rudder .... As the rudder moves along the groove, its change in orientation would bring about a change in the orientation of the front wheels. Because the shape of the groove is isomorphic with [i.e., resembles] the curve itself, the wheels change along with the S-curve and the vehicle moves through it without ever bumping into a wall. (Ramsey, 2007, p. 198)

So, in order to succeed at getting through the track, the car must move in a way that is adapted to external conditions, namely to the shape of the track. To achieve this, it uses a simple mechanism comprising of: (1) the wheels, (2) the steering wheel, (3) the groove (internal map), and (4) the rudder, which establishes a causal connection between the groove and the steering wheel. The function of the steering wheel is to move the wheels (the car's "effectors") in a way that is adapted to the shape of the track. Thus, the shape of the track constitutes conditions of success for the steering wheel's activity in the mechanism. At the same time, the steering wheel is not directly causally coupled with the track. Instead, the whole mechanism works in such a way that the steering wheel makes use of the representational vehicle – the groove – to succeed at performing its function. The steering wheel benefits functionally from the groove, and in this sense the groove's function can be described as "action guidance": it indirectly guides the car (the system as a whole) by directly "guiding" the steering wheel (internal consumer). Therefore, from the point of view of the present proposal, the track itself is the represented object, the groove plays the role of representational vehicle, and the steering wheel is the representation consumer.

This simple example also illustrates nicely what I mean when I say that, in a representational mechanism, structural resemblance between the representational vehicle and the represented object is *relevant* to the way the vehicle functions (see also Ramsey, 2007). Notice that the car described above cannot achieve navigational success unless its internal map (the groove) spatially-structurally resembles the spatial structure of the track; the car's success *depends* on this relation. So, in a representational mechanism, structural resemblance is relevant in the sense of being both *causally* and *explanatorily* relevant. It is causally relevant because by intervening in whether the resemblance holds between the vehicle and the represented object – for example, by alternating the track's or the groove's shape (or both) in the car example – we can manipulate the consumer's success in performing its function (see Woodward, 2003). It is explanatorily relevant because if we are dealing with a representational mechanism, any explanation of a given phenomenon – like the car's ability to get through the track successfully – will be hopelessly incomplete unless it mentions how the mechanism in question exploits the resemblance between the vehicle and the representational object (see also Ramsey, 2007).

3.4. Decouplability condition

One might wonder whether the two conditions discussed so far already provide us with a notion of representation robust enough to meet the JDC. Mechanistically realized, action-guiding s-representations seem to function in a way that is very analogous to the way external s-representations do. For example, an external map also provides its user with action guidance by enabling her to exploit the structural similarity between the map itself and the terrain it represents (the success of map-using practice depends on whether the map structurally corresponds to the terrain). We should remain careful, though, because it might be argued that this way of understanding representations is still open to counterexamples – cases, where the two abovementioned conditions are met, but nonetheless we are dealing with a clearly nonrepresentational structure. For example, we would not characterize a key as representing a lock simply because our success in opening the door depends on whether the key's shape resembles the lock's shape (Tonneau, 2012). Perhaps in order to make sure that our notion of representation meets the JDC, some additional constraints on representational mechanisms are needed. This is why I postulate two additional conditions that need to be met in order to earn a mechanism representational status.

My third condition appeals to the familiar idea that representations should be decouplable from the states of affairs that they represent. Representations are characterized by the fact that they can play their representational role even when what is represented is not "reliably present and manifest" for the representation-using system or mechanism (Haugeland, 1998). If this is right, then any truly representational system should have the ability to use representations to guide its actions even when the representational object is not present (see Chemero, 2009; Clark & Grush, 1999; Grush, 1997). In other words, we are not dealing with a truly representational system or mechanism unless the structures we want to treat as representations can be used off-line, outside of direct interactions with the representational object. For example, representations should afford their user to guide her/its actions (plan them, make decisions about them, etc.) with respect to states of affairs that are merely possible or located in the future.

Decouplability constitutes my third condition of being a representational mechanism. That is, it should be possible for a representational vehicle in a representational mechanism to perform its function off-line. I do not mean to exclude off-hand the very possibility of using representations on-line, in a way that is directly coupled with environmental conditions. Nonetheless, to earn the status of a representational vehicle, a component part of a mechanism should at least be *able* to perform its action-guiding function in a decoupled way.

Given that the very notion of decouplability is not quite clear and can be interpreted in different ways (Chemero, 2009), it is important to clarify it and cash it out in mechanistic terms. According to the view I am advocating, whether the representation is decouplable depends on the causal position that either the vehicle-consumer configuration or the mechanism as a whole occupies with respect to the representational object. There are two levels of decouplability: weak and strong decouplability. It is then a *minimal* condition of being a representational mechanism to be a mechanism that uses representations in a weakly decouplable way.

Here is how I understand weak and strong decouplability, with a simple illustration for each:

i. An internal representation is *weakly decouplable* iff it is possible for it to perform its action-guiding function when (1) there is no causal connection between the representational vehicle and the representational object, and (2) there is no causal connection between the representation consumer and the representational object.

An example of weak decouplability: The car example presented in the previous subsection provides a nice illustration. Notice that, in this case, there are no causal interactions between the track – the representational object – and either the groove (the vehicle) or the steering wheel (the consumer). In this particular sense, the groove stands in for the track. Nonetheless, this is not a case of strong decouplability because the internal map of the track is only useful for the larger mechanism when the car as a whole is interacting with the track. So the represented object always *is* reliably present for (viz. causally connected to) the car as a whole.

ii. An internal representation is *strongly decouplable* iff it is possible for it to perform its action-guiding function if there is no causal connection between the representational object and the whole system or mechanism comprising of (among others) the representational vehicle and the representation consumer.

An example of strong decouplability: Imagine a modified version of the car discussed above. This one does not simply use a map of the track, but is also equipped with a *miniature version of itself moving through a miniature version of the track.* Let us also imagine that this miniature car (1) can be placed in different variations of a miniature track (differing in shape) and (2) can be equipped with different variations of a miniature internal map (differing in the shape of the groove). This manipulable miniature enables the car to simulate different possible courses of action – corresponding to alternative track-map combinations – *before* it even starts moving through the large track. In other words, this sort of process of miniature-use enables the car to preselect one among many possible interactions with its environment. For example, we might imagine that thanks to this sort of internal configuration, the car can test multiple variations of the map in order to select one that is best suited to successfully drive it through a track with a particular shape (say, an S-shaped track).[9] In such a case, the consumption of representation is postponed until the car actually starts driving through the proper (large) track. What we are dealing with here is a situation where the *larger car as a whole* has at its disposal an internal, mechanical model of itself that enables it to plan interactions with the world *before* it actually engages in them. This sort of internal representation enables the system to guide its action with respect to states of affairs that are not "reliably present" for the system because they are not present for the system *at all.* For this reason, this is an example of a strongly decouplable representation.

3.5. Error detection condition

My fourth and last condition once again has its origins in Bickhard's interactivist theory of representation (Bickhard, 1999, 2004; see also Anderson & Rosenberg, 2008; Miłkowski, 2013). Bickhard proposed that a theory of representation should not only account for the possibility of representational error, but also account for how representational error is *system-detectable.* That is, a good theory of representation should have the conceptual resources to show how a system or mechanism using the representation is able to detect situations in which the representation is false or inaccurate.

Now, as Bickhard argues, it is impossible to account for system-detectable error if one assumes that representations are constituted by a correspondence – grounded, for example, in covariance or evolutionary history – between the representation itself and what is represented. This is because no organism can, so to speak, look from the "outside" to gain cognitive access to whether the appropriate correspondence-establishing relation actually holds (see Bickhard, 1999, 2004 for a more detailed version of this argument). Bickhard argues that the ability to recognize representational error can only be explained if one's theory closely connects representations with (inter)action. To make a long story short, on his account, representational content is determined by the "dynamic presuppositions" of an action that is based on a given representational state; that is, content is determined by conditions that should occur if a given representation-guided action is to be successful (see also section 3.2). Representational error occurs whenever there is a mismatch between the action based on a representation and environmental conditions, that is, whenever action's conditions of success do *not* occur. So on this approach, in order to detect representational error, the representation user does *not* need to reach out to the world itself to check whether it matches the representation. Rather, representational error can be detected through the detection of *action failure*, which seems to be something easily accessible for the acting organism. Failure of action indicates that the dynamical presuppositions were false, and so the representation in use was false.

Borrowing from Bickhard, I want to make the possibility of error detection the fourth and last condition of being a representational mechanism. To be genuinely representational, a mechanism needs internal resources that enable it to detect situations when the representation used by the consumer is false.[10]

Of course, to account for false representation detection, one needs first to account for false representation as such, and this in turn requires some theory of content or, in other words, of how the representation's accuracy or truth conditions are determined. Although in the present article I am concerned with representational function and not representational content – with what it is to function as a representation rather than with how it is determined what the representation is about (see Ramsey, 2007, 2015) – let me briefly sketch what content might be on the view I am advocating. Whenever the representation consumer's activity is guided by the representational vehicle, there are two senses in which a representational "object" is involved. A representational object in the first sense is anything with respect to which the representation-guided action is *de facto* performed. A representational object in the second sense is that with respect to which the representation-guided action should (*de iure*) be performed in order to be successful. The self-driving car example may serve as an illustration once again. Imagine that the car drives through an S-shaped track, but it uses a Ƨ-shaped map, so the car crashes and fails to get through the track. In this case, the S-shaped track is a representational object in the first sense: that which the representation is "applied" to, so to speak. An *Z-shaped* track constitutes the representational object in the second sense: it is the track that determines the representation's *conditions of successful action guidance* (that is, conditions that should occur if the vehicle is to guarantee the success of the consumer). It is the representational object in this latter sense which we should, I think, equate with representational content. If this is so, then representational error occurs whenever there is a discrepancy between representational objects in those two senses; that is, whenever the actual environmental conditions that the representation-guided action *is* performed with respect to do not match the conditions it *should* be performed with respect to in order to be successful.[11]

Given this sort of account of content, error detection can be dealt with along Bickhardian lines. Namely, representational error detection can be achieved in a representational mechanism if the mechanism has the ability to detect *action* failure. That is, the mechanism in question should be equipped with internal components whose function is to detect[12] the fact that the action guided by the representational vehicle (through its effect on the representation consumer) fails to achieve success. This way, the mechanism has the ability to "recognize" something that is *not* directly accessible for it – namely, the fact that the representational vehicle does not match (structurally resemble) the world – by recognizing something that *is* directly accessible for it, namely the fact that the action based on the representation has failed.

How can the ability to detect representational error be realized in a mechanism? It seems that there are at least two ways of setting up a mechanism so that it achieves error detection. I will call those the "action-failure-detection" strategy and the "predict-and-compare" strategy. Both strategies require the existence of a component part that provides the mechanism with a feedback signal from the world, but they differ in how this signal is used in the mechanism. Let me discuss both strategies in turn.

i. *Action-failure-detection* strategy: The mechanism is equipped with a feedback channel that monitors the action (e.g. through monitoring the body) and detects situations where the feedback signal indicates an action failure, and thus a false representation.

An example of action-failure detection strategy: Consider the sophisticated version of the self-driving car, the one that uses a miniature model of itself. Suppose the car has performed a number of internal simulations and selected an internal map that has been predicted to be the most successful when it comes to the drive through the proper track. Now, imagine that when the car starts driving, it bounces off the sides of the track a couple times and eventually fails to get through. This lack of success can be used as an indication of the fact the map in use – the one preselected using the internal simulation – is inaccurate. Imagine that the car has a set of sensors on each side and a feedback channel, such that a feedback signal is generated and sent through this channel whenever the car touches (bounces off) the side of the track. The feedback signal is fed to a detector that gets activated when a certain number of bounces occurs. Upon its activation, the detector stops the car and initiates the internal simulation once again, in order to select another map which resembles the shape of the track more closely (see also note 9).

ii. *Predict-and-compare* strategy: The mechanism is equipped with a feedback channel that monitors the action (e.g. through monitoring the body). The representational vehicle generates an internal "mock" or predictive signal that (1) systematically depends on the state that the vehicle is in, (2) matches the feedback signal that would be received if the action guided by the vehicle was to be successful (viz. if the representation being used was true). Also, the mechanism has a comparator part that compares the predictive signal with that actual feedback signal and measures the discrepancy between the two, such that when the discrepancy reaches a certain level, it is used as an indication of an action failure, and thus of a false representation.

An example of predict-and-compare strategy: Let me modify the sophisticated self-driving car example once again. Imagine that the car has infrared sensors on each side that systematically co-vary with the distance between a given side of the car and a given side of the track. Imagine also that the internal miniature model of the car has the same type of sensors that react to its distance from the sides of the model track. As the model car drives through the model track, the activity of its sensors is registered and tagged as corresponding to a specific map-track combination. Now, when the large car starts driving through the proper track, its sensor activity sends a feedback signal to the mechanism. The car has a component part that compares this feedback signal against the signal predicted in the simulation mode. When there is a discrepancy between the two, or if the discrepancy reaches a certain level, this indicates a mismatch between the simulation and the actual action. The comparator stops the car and initiates the internal simulation once again, in order to select another map which resembles the shape of the track more closely. Here is a diagram showing how a representational mechanism using the predict-and-compare strategy is organized:

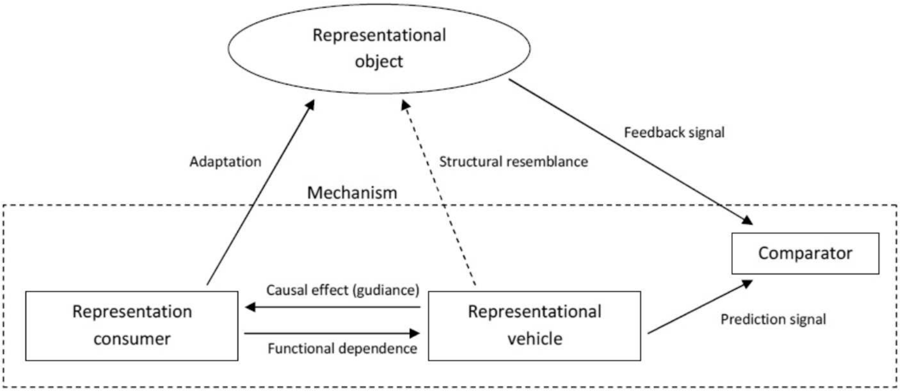

Figure 2. A schematic diagram of a representational mechanism equipped with a predict-and-compare system for representational error detection (see main text for details)

4. Representational mechanisms: addressing possible worries

Before closing, let me address two possible worries about the account of representational explanation that I have laid out in the previous section:

i. How do we know that the present account of representational explanation meets Ramsey's JDC?

As has been mentioned in section 2, I follow Ramsey (2007) in assuming that for an explanation to be truly representational, it should be possible to show how states or structures postulated as representations in this explanation actually play genuinely representational functional roles. This is the JDC. I think that the major strength of the present proposal lies in the fact that it views representations in a way that meets Ramsey's challenge. The theory of representational mechanisms on offer here shows conditions that need to be met if an explanation of a given phenomenon is to be judged as truly, nontrivially, and in an illuminating way representational; or, in other words, as an explanation in which the notion of representation does not serve as a "filler term" (Craver, 2007), but actually has some specific and important explanatory job to do. However, someone might demand that I be more explicit as to why *exactly* I claim that representations as I view them meet the JDC. How do we know that structures which meet the four conditions discussed above truly function as representations in cognitive mechanisms?

Let me start by taking a closer look at the method of examining the explanatory status of different notions of representation that Ramsey himself (2007) employs in his book. According to this method, in order to demonstrate that a given structure functions as a representation (and thus meets the JDC), one needs to show that the way this structure functions is to some nontrivial extent analogous to the way structures that we pretheoretically categorize as representations function. I want to use this strategy here. I think that what I categorize as *internal* representations in my theory deserve this status because the role they play in a mechanism sufficiently resembles the role played by *external* s-representations.

Take an example of an external representational artifact like a map. What is it about the functioning of a map that makes it a representation? I think the following characteristics are important. First, the map (which constitutes a representational vehicle) structurally resembles the terrain it is a map of (the representational object): the metric structure of the map constituents mirrors – albeit in an incomplete, idealized fashion that ignores a large amount of details – the metric structure of the represented terrain. This resemblance relation does not simply hold between the map and the terrain, but is relevant to the functioning of the map; the map is useful to the extent that it actually resembles the terrain. Second, the map serves as an action guide for its human user (representation consumer), by virtue of enabling the user to navigate the terrain it represents, to plan a route to a given destination, to draw conclusions about relative distances between the elements of the terrain, etc. Third, the map can perform its action-guiding function even when its user is not engaged in direct interactions with the represented terrain. For example, it can be used for off-line planning, making inferences or even pondering alternative routes purely counterfactually. Fourth, because of how the map is involved in action-guidance, it affords its user to recognize its inaccuracies as a representation. The user can detect these inaccuracies if the actions guided by the map fail to achieve their goals: when the user fails to reach a given point despite apparently taking the proper route, or if it takes longer to reach a given point than it should take according to the map, etc.

What characterizes maps, then, is that they meet all four conditions discussed in section 3: they are decouplable, action-guiding, error-detection-affording structural representations. Now, the idea behind my proposal about representational mechanisms is that mechanisms of this kind use *internal functional analogs* of external, artifactual representational devices like maps. The only difference lies in the fact that those former representations work in a purely "mechanical" manner; that is, the process of their consumption does not require the consumer to possess full-blown interpretational abilities characteristic of human intentional subjects (see also Ramsey, 2007). Internal representations, as they figure in cognitive-scientific, mechanistic explanations, fulfill the conditions mentioned in section 3 not because they can be interpreted by humans, but because of their place in a functional-causal architecture of purely subpersonal mechanisms.

ii. Is the present account of representation too restrictive?

Ramsey (2007) puts forward a diagnosis according to which most notions of representation routinely employed by cognitive scientists are too liberal. That is, the criteria that cognitive scientists use to categorize something as a representation are so inclusive that *too many* things get to be labeled as "representations". One might wonder if my proposal does not suffer from an opposite problem. Isn't my account of representational explanation in cognitive science *too restrictive*? Doesn't my account render too many cognitive-scientific theories, models, and explanations *non*representational? Are there even any examples of existing cognitive-scientific theories, models or explanations that *do* count as representational according to my proposed criteria?

There are two ways to address this worry. One is simply to bite the bullet and admit that the account on offer here *is* restrictive. This is simply the price to pay if one is to make sure that the structures one calls "representations" meet the JDC, and consequently that the (mechanistic) explanations these structures feature in are truly representational explanations. Another way to address the restrictiveness worry is to point to the fact that there *are* at least two theories in cognitive science that meet the criteria put forward in section 3. So my proposal is not unrealistic or detached from the actual explanatory practice of cognitive scientists. Space forbids me from discussing these two examples in detail. Doing this would require a separate article. So let me just name the examples and present them in a sketchy and preliminary fashion.

The first example is Rick Grush's (1997, 2004) famous emulation theory, according to which motor control, perception, and imagery make use of internal emulators of the body and the world. I think that Grush's theory meets my four conditions in the following ways. First, the structural resemblance condition is met because of how emulators owe their functioning to the way their activity dynamically mimics the dynamics of the body, or the body-world loop. Second, the action-guidance condition is met in virtue of the fact that emulators enable their consumers (like the motor system) to practically orient with respect to the world, especially when it comes to precise, anticipatory movement control. Third, the decouplability criterion is met because of how emulators, according to Grush (2004), can be used off-line, to subserve imagery, planning, or even counterfactual reasoning. It is possible, then, to use emulators in a strongly decoupled way. Fourth, emulator-based sensory predictions are, through the use of Kalman filtering, continuously compared against the feedback signal coming from the body and the world. This enables the emulator-using mechanism to use the predict-and-compare strategy to detect discrepancies between what has been predicted to happen using the representation (the emulator) and what is actually happening.

The second example is the predictive coding theory, which is now gaining momentum in cognitive science and which assumes that perception and action are subserved by (purely mechanically or subpersonally realized) Bayesian reasoning which enables the brain to infer hidden worldly causes of incoming sensory signals (see Clark, 2013; Huang & Rao, 2011). Admittedly, in this case there are some interpretational challenges, but let me present the gist of how I think the predictive coding framework might count as representational in light of my proposal.[13] First, the structural resemblance condition is met due to the fact that according to the theory, the brain constructs internal generative models, in which the structure of hidden or latent variables corresponds to or mirrors the causal structure of the world. This similarity in structure is relevant because the predictive accuracy of a generative model depends on it. Second, the action-guidance condition is met in virtue of the fact that generative models are constructed and used in order to control action and to enable perceptual categorization. Third, assuming that the top-down production of "mock" sensory signals by generative models temporally precedes the predicted feedback signal from the world, generative models can be described as at least weakly decouplable. It is possible that they are also strongly decouplable, given the fact that the theory explains imagery as a result of top-down, generative-model-based sensory signal production (see Clark, 2013). Fourth, the error-detection condition is met because, according to the predictive coding theory, the cognitive system (brain) is continuously engaged in prediction error minimization. That is, the bottom-up feedback signal from the world is constantly compared against the top-down signal generated by the generative model, so that the system can recognize if the model is in error and, if so, modify the model-based hypothesis to one that better corresponds to the world.

5. Conclusion

The aim of this paper has been to propose a solution to what I have called the "second-order problem" of the explanatory role of internal representations in cognitive science. That is, instead of arguing for either representationalism or antirepresentationalism, my intention was to answer the question of how we should construe the very nature of representational explanations in cognitive science. I have assumed that the explanations in question are mechanistic, i.e. explain the phenomena by their underlying mechanisms, understood as a sets of organized, active component parts. The question was, then: What makes a mechanistic explanation representational? According to my proposal, representational mechanisms are ones that make use of decouplable, action-guiding, error-detection-affording structural representations of (some aspects of) the world. A mechanistic explanation of a given cognitive phenomenon is representational if the mechanism it postulates as underlying this phenomenon is representational.

Obviously, the proposal on offer here does not solve the first-order problem of representation: it tells us what representational explanations are (or should be), but by itself it remains silent about whether we can actually explain any phenomena – and if so, which ones – representationally. However, it should be clear that the theory of representational mechanisms can serve as an important step towards resolving the first-order problem. After all, it provides us with criteria that need to be met if representationalism is to be proclaimed – in a justified way – the winner of the representationalism/antirepresentationalism debate.

NOTES

[1] My assumption is not that the only way of explaining phenomena available for a cognitive scientist is describing their mechanisms. For example, evolutionarily-oriented cognitive scientists often explain phenomena not by their proximate mechanisms, but by their ultimate, evolutionary causal etiology.

[2] Within the mechanistic framework, functions should be understood as Robert Cummins' "role functions" (Cummins, 1975; Craver, 2001, 2007). That is, the function of a component of a mechanism of a given phenomenon is determined by the way this component contributes to producing this phenomenon. For example, the heart's function as

a component of a larger mechanism of blood distribution in the organism is to pump blood and not to make rhythmic noises, since the heart contributes to blood distribution by virtue of the former, and not the latter activity.

[3] Notice that structures in question need not be of the same kind. For example, in weather maps – which constitute an example of an external representation based on structural resemblance – *spatial* structure (spacing of so called "isobars") is used to represent *atmospheric* structure of pressure gradients (O'Brien & Opie, 2004). This opens up the possibility that *worldly* structures (spatial, temporal, categorical, social, etc.) might be represented in *neural/computational* structures of the brain.

[4] This idea might be used to deal with other arguments raised against similarity-based accounts of representation. Consider two such arguments. According to one of them, representations cannot be grounded in structural similarity because similarity is reflexive and symmetrical, and the representational relation is not. Notice, however, that this problem arises only for the view that the existence of resemblance is sufficient to make something a representation. This is not the claim I am making here. What I claim is that to be a representation, a component part of a mechanism must be such that its proper functioning depends on whether it structurally resembles something else. But the relation of x being functionally dependent on structural similarity between x itself and some other y is neither reflexive, nor symmetrical (see a similar solution in Bartles, 2006). According to another argument, structural similarities come cheap and are prevalent in the world (see Bartles, 2006; Ramsey, 2007). If so, then the grounding of representation in structural resemblance might lead to panrepresentationalism, which would trivialize the notion of representation. However, even if structural similarities are cheap, functional, or exploitable, structural similarities are not. And it is this latter kind of similarities that my theory pertains to.

[5] What do I mean by conditions of success? Discussing this subject at length is beyond the scope of the present article, so let me simply follow Bickhard (1999, 2004) and say that "success" here depends on whether a given mechanism contributes to keeping the larger system (organism) in a far from thermodynamic equilibrium state. So, in our case, R1 enables the organism to minimize entropy in conditions R1', and R2 enables the organism to minimize entropy in R2'.

[6] Despite the fact that I am using an example of purely motor action-guidance, my proposal should not be read as restricting the role of representations to the motor domain. My proposal does not preclude representations from guiding cognitive actions, e.g. decision making, mental state attribution, or categorization (see also Anderson & Rosenberg, 2008).

[7] In this specific sense, they are subpersonal representations.

[8] The major difference between how I and Millikan understand the representation consumer lies in the difference regarding how we perceive the nature of functions as such. According to Millikan (1984), a structure functions as a consumer because of its evolutionary history. According to the view on offer here, consumers owe their status to their Cummins-style role function in a larger mechanism, viz. to the fact that they causally contribute to the larger mechanism in a way that functionally depends on the representational vehicle (see also note 2).

[9] We might assume that the car starts by randomly generating different miniature-map-variations, and then performs a number of simulated drives through a miniature S-shaped track, with each simulation using a different version of a miniature map. Then the most successful miniature map is selected and used as template to create an internal map that the larger car will actually use to navigate its way through the large track. For an actual example of a somewhat similar mechanism from robotics, see Bongard, Zykov and Lipson, 2006.

[10] For a previous attempt at marrying the idea of system-detectable representational error with the mechanistic framework, see Miłkowski, 2013.

[11] If my theory treated content as being simply determined by the structural resemblance relation, it would belong in the category of correspondence-based theories of representation and thus fall prey to problems that all theories of this sort face when it comes to explaining system-detectable error. But by my account, content is not determined solely by the structural resemblance relation. Notice, however, that action-guidance success conditions – which do determine content – actually occur if there is a structural resemblance between the vehicle and the representational object in the second sense (for example, if the shape of the groove resembles the shape of the track that the car is driving though). So the existence of this resemblance guarantees that the representation is true or accurate; and lack of therefore means that the representation is false or inaccurate.

[12] Here, I understand "detection" in a completely non-representational way. That is, I do not mean to imply that detecting representational error requires the mechanism to represent the fact that it is using a false representation. Some sort of non-representational reliable indicator that makes the mechanism reactive to situations in which it uses a false representation will suffice. This way, I avoid assuming representations to explain representations (see Bickhard, 1999, 2004).

[13] I am assuming that predictive coding framework can and eventually will be supplemented with neural-level details and so count as providing full-blown mechanistic explanations of phenomena (see Jones & Love, 2011).

REFERENCES

Anderson, M. L., & Rosenberg, G. (2008). Content and action: The guidance theory of representation. *Journal of Mind and Behavior*, *29*, 55–86.

Bartels, A. (2006). Defending the structural concept of representation. *Theoria*, *55*, 7–19.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanistic alternative. *Studies in History and Philosophy of the Biological and Biomedical Sciences*, *36*, 421–441.

Bickhard, M. H. (1999). Interaction and representation. *Theory and Psychology*, *9*, 435–458.

Bickhard, M. H. (2004). The dynamic emergence of representation. In H. Clapin, P. Staines & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 71–90). Oxford: Elsevier Science.

Bongard, J., Zykov, V., & Lipson, H. (2006). Resilient machines through continuous self-modeling. *Science*, *314*, 1118–1121.

Calvo Garzon, P. (2008). Towards a general theory of antirepresentationalism. *British Journal of Philosophy of Science*, *59*, 259–292.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Clark, A., & Grush, R. (1999). Towards a cognitive robotics. *Adaptive Behavior*, *7*, 5–16.

Craver, C. F. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*, 31–55.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*, 741–765.

Cummins, R. (1989). *Meaning and mental representation*. Cambridge, MA: MIT Press.

Cummins, R. (1996). *Representations, targets and attitudes*. Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Fodor, J. (1994). *The elm and the expert: Mentalese and its semantics*. Cambridge, MA: MIT Press.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*, 342–353.

Gładziejewski, P. (2015). Action guidance is not enough; representations need correspondence too: A plea for a two-factor theory of representation. *New Ideas in Psychology*. doi:10.1016/j.newideapsych.2015.01.005.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*, 5–23.

Grush, R. (2004). The emulation theory of representation: motor control, imagery and perception. *Behavioral and Brain Sciences*, *27*, 377–442.

Haugeland, J. (1998). Representational genera. In J. Haugeland (Ed.), *Having thought: Essays in the metaphysics of mind* (pp. 171–206). Cambridge, MA: Harvard University Press.

Haselager, P., de Groot, A., & van Rappard, H. (2003). Representationalism vs. anti-representationalism: A debate for the sake of appearance. *Philosophical Psychology*, *16*, 5–23.

Huang, Y., & Rao, R. (2011) Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 580–593.

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

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contribution of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 188–231.

Machamer, P., Darden, L., & Craver, C. (2001). Thinking about mechanisms. *Philosophy of Science*, *67*, 1–25.

McDowell, J. (1994). The content of perceptual experience. *Philosophical Quarterly*, *44*, 190–205.

Millikan, R.G. (1984). *Language, thought and other biological categories.* Cambridge: Cambridge University Press.

Miłkowski, M. (2013). *Explaining the computational mind.* Cambridge, MA: MIT Press.

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 1–20). Oxford: Elsevier Science.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese, 183*, 283–311.

Ramsey, W. (2007). *Representation reconsidered.* Cambridge: Cambridge University Press.

Ramsey, W. (2015). Untangling two questions about representation. *New Ideas in Psychology.* doi:10.1016/j.newideapsych.2015.01.004.

Shea, N. (2007). Consumers need information: Supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research*, *75*, 404–435.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, *87*, 449–508.

Tonneau, F. (2012). Metaphor and truth: A review of *Representation reconsidered* by W. M. Ramsey. *Behavior and Philosophy*, *39/40*, 331–343.

Woodward, J. (2003). *Making things happen: A theory of causal explanation.* Oxford: Oxford University Press.