

Manuscript accepted. To appear in *Synthese*

Link: <https://link.springer.com/article/10.1007/s11229-019-02505-1>

Similarity-Based Cognition:

Radical Enactivism meets Cognitive Neuroscience

Miguel Segundo-Ortin

School of Liberal Arts
Faculty of Law, Humanities and the Arts
University of Wollongong
NSW 2522, Australia
Email: miguel.segundo.ortin@gmail.com
ORCID: 0000-0002-3787-0013

Daniel D. Hutto

School of Liberal Arts
Faculty of Law, Humanities and the Arts
University of Wollongong
NSW 2522, Australia
Tel: +61 2 4221 3987
Email: ddhutto@uow.edu.au
ORCID: 0000-0002-3585-4618

Abstract: Similarity-based cognition is commonplace. It occurs whenever an agent or system exploits the similarities that hold between two or more items — e.g., events, processes, objects, and so on— in order to perform some cognitive task. This kind of cognition is of special interest to cognitive neuroscientists. This paper explicates how similarity-based cognition can be understood through the lens of radical enactivism and why doing so has advantages over its representationalist rival, which posits the existence of structural representations or S-representations. Specifically, it is argued that there are problems both with accounting for the content of S-representations and with understanding how neurally-based structural similarities can work as representations (even if contentless) in guiding intelligent behavior. Finally, with these clarifications in place, it is revealed how radically enactivism can commit to an account of similarity-based cognition in its understanding of neurodynamics.

Keywords: similarity-based cognition; cognitive neuroscience; radical enactivism; S-representations; job description challenge; hard problem of content.

Introduction

Similarity-based cognition occurs whenever an agent or system exploits the similarities that hold between two or more items — e.g., events, processes, objects, and so on— in order to perform some cognitive task. It has been proposed that conceiving of representations in terms of structural similarities provides new resources to overcome traditional puzzles that have plagued

other representational theories of cognition. With respect to the explanatory needs of cognitive neuroscience, many philosophers maintain that appealing to similarity-based forms of cognition is the most promising strategy for building an adequate representational theory of mind (O'Brien 2015a, 2015b; Gladziejewski and Miłkowski 2017; Williams 2017; Ramsey 2018; Shea 2018; Lee 2018).

The purpose of this paper is two-fold. First, it challenges the two main strategies that motivate conceiving neurally-based structural similarities of interest to cognitive neuroscience as mental representations. Second, once this clarification is made, the paper explicates how neurally-based structural similarities can play a part in explaining intelligent behaviour within a radically enactive and embodied account of cognition.

The structure of this paper is as follows. Section 1 introduces the notion of similarity-based cognition, showing how it is cashed out in cognitive neuroscience. It focuses on the well-known example of place cells in rats' hippocampus. Section 2 introduces the arguments in favor of the idea that neurally-based structural similarities should be conceived of as contentful structural or S-representations. Section 3 problematizes those arguments, showing that we have yet to be supplied reasons for thinking that structural similarities – those that purportedly do explanatory work in cognitive neuroscience – are contentful. It is argued that defenders of S-representational interpretation of structural similarities presuppose, but do not explain, the origin of the contentful properties of structural similarities. If this analysis holds good, despite claims to the contrary, S-representations fail to answer the job description challenge. Section 4 examines a less demanding reason for thinking that neurally-based structural similarities should be conceived of as S-representations. Putting aside questions about content, some argue that the

alleged fact that structural similarities are exploited by brains in a way that is analogous to the ways cognitive agents exploit models or maps suffices to establish the representational status of structural similarities. New empirical findings concerning the future-oriented activity of place cells are called upon to justify this position. Against this view, it is argued that these empirical findings provide no support for the claim that place cells are used in anything like route-planning or surrogate reasoning. It follows that if there is no robust, non-metaphorical sense in which the brain uses structural similarities as models or maps this inferential path to the conclusion that structural similarities have representational status is blocked. Finally, section 5 explicates how similarity-based cognition can be understood under the auspices of a radical enactive, non-representational conception of cognition.

1. Similarity-Based Cognition in Cognitive Neuroscience

Similarity-based cognition, SBC, is commonplace. It occurs whenever an agent exploits relations of similarity holding between two or more items — e.g., events, processes, objects, and so on — in order to perform a cognitive task (Cummins 1994; Godfrey-Smith 2009; Ramsey 2018).

Everyday examples of SBC include navigating to a location by using a cartographic map or using a mercury thermometer to discover the current temperature of a room. In such cases, similarities between items are exploited by cognitive agents when they treat one item, X, as a surrogate or stand in for a target item, Y, and when doing so reliably guides behaviour with respect to the target item.¹ Behaviour is understood broadly here as to include an agent making embodied responses, taking particular actions, making reliable judgements or inferences, and explicitly deciding on courses of actions.

Many philosophers and cognitive neuroscientists assume that fundamental varieties of cognition are similarity-based (Shagrir 2012; Jacobson 2015; Kriegeskorte and Kievit 2013; Sachs 2018; Gladziejewski and Miłkowski 2017, Williams and Collings 2017; Shea 2018). The guiding hypothesis, when SBC is pitched at the neural level, is that neural states make a causal contribution to the processes that drive the behaviour of cognitive systems towards some target items in virtue of the fact that certain similarities hold between the neural states and those target items. Call this “neurally-based SBC.”

A thing can be similar to another in many ways, nonetheless. One item may be similar to another simply by having shared colour, mass, charge, and so on. However, neural items do not enable cognitive work to get done by possessing just any kind of similarities with their distal targets. Rather, the similarities at issue are thought to be of a second-order or structural kind.

O’Brien and Opie (2004) explicate what structural similarity means in the following terms:

We will say that one system *structurally resembles* another when the physical relations among the objects that comprise the first preserve some aspects of the relational organization of the objects that comprise the second. (pp. 14-15, emphasis original)

Cartographic maps, for example, rely on structural similarities. A cartographic map of Sydney can help us to get around a specific location of the city just in case its relevant constituents (lines, figures, symbols, and so on) are arranged in a way that systematically mirrors the topographic and metrical relations that hold among the relevant constituents (buildings, streets, and so on) of the city of Sydney. This structure-preserving mapping relationship holding between the constituents of both systems, the map, and the city,

guarantees that specific points on the map correspond to specific locations in the city. The same reasoning applies to mercury thermometers. Mercury thermometers inform us about the temperature of a particular room just in case the variations in the height of the liquid column correspond to orderly variations in the temperature of that room.

When applied to the domain of cognitive neuroscience, the idea is that neural items can play a causal part in enabling the achievement of a cognitive task in virtue of the fact such neural items mirror relevant structural properties of some target domain.² Accordingly, whether the cognitive task can be successfully achieved depends, at least in part, on the degree to which the neural item at issue structurally mirrors the relevant target domain.

For cognitive neuroscientists, neurally-based SBC can be called upon to explain a vast array of cognitive phenomena – including visual perception, motor control, memory, imagination. In recent years, the remarkable potential of such explanations has become evident with the advent of predictive processing accounts of cognition – those which assume that the best and most unified explanations of cognitive phenomena need to posit generative models operating across multiple scales and levels (Clark 2016; Gładziejewski 2016; Williams 2017; Sachs 2018).

Neurally-based SBC is most famously invoked in explanations of rodent spatial navigation (see Bechtel 2016; Shea 2018). Years after Tolman hypothesized the existence of a “cognitive map” to explain spatial navigation in mammals (Tolman 1948), researchers discovered that individual neurons allocated into regions DG, CA1 and CA3 of the rats’ hippocampus fire “solely or maximally when the rat was situated in a particular part of the testing platform facing in a particular direction” (O’Keefe and Dostrovsky 1971, p. 172). Follow-up studies by O’Keefe (1976) and O’Keefe and Nadel (1978) also

showed that neurons in the CA1 region respond primarily to location, with some of them—20 out of the 26 firstly reported—firing when the rats occupy or run past the appropriate location. Scientists called the hippocampal neurons involved in navigation tasks “place cells.”

The current scientific consensus assumes that place cells within the rats’ hippocampus play a causal role in enabling navigation tasks — namely, finding shortcuts towards a food source — because they are structurally similar to the environment. More precisely, it has been discovered that the patterns of co-activation relationships between the cells (roughly, their tendency to show joint activity) mirror the structure of metric relations among different relevant locations within the environment (Bechtel 2016; Knierim 2015; Pfeiffer and Foster 2013). Accordingly, the structure-preserving mapping relationship holding between the activation profile of the place cells and the spatial layout of the environment contributes to explaining the capability of rats to successfully navigate the environment, even in complete darkness.

Given this, place cells are taken to be the realizers of the cognitive map previously hypothesized by Tolman (Schmidt and Redish, 2013). For example, it has been claimed that the whole system of place cells provides other parts of the rats’ brains with “an internal map of the spatial layout of the environment, encoded in a Cartesian coordinate system” (Gładziejewski and Miłkowski 2017, p. 344; O’Keefe and Dostrovsky 1971, p. 174).³ In the same vein, Ramsey (2016) holds that place cells function “as component parts of an encoded map of the environment that the rat is trying to navigate. They ... are serving as surrogate stand-ins within a broader map-like neural structure” (p. 9).

2. The Case for Structural Representations

For many, the very existence of causally potent structural similarities at the neural level entails the existence of a special kind of mental representation. Proponents of this idea call these mental representations “structural” or “S-representations.”

As canonically described, S-representations can be conceived of as components of larger cognitive mechanisms – where such mechanisms are partly individuated by reference to the function they perform. Imagine a mechanism M which is responsible for cognitive function F . M is an S-representational system if its ability to perform F causally depends on the fact that at least one of its components R is structurally similar to some target domain T .⁴ If that is the case, M can fail to perform F if R is not sufficiently structurally similar to T ; and, analogously, if M succeeds in bringing about F , it is in part due to the fact that R is sufficiently structurally similar to T .⁵

Nonetheless, despite the popularity and promise of S-representations in cognitive neuroscience, if such neural items are to count as bona fide cognitive representations, they must face up to the “job description challenge” (Ramsey 2007).⁶ According to this challenge, if something is to count as a cognitive representation it must satisfy two conditions.

First, it has to be shown that the structure in question possesses content such that it refers, denotes or depicts something else as being a certain way.⁷ Canonically, a cognitive state or structure bears representational content if and only if it has conditions of satisfaction of a special sort – namely if it is susceptible to being true or false, veridical or non-veridical, right or wrong, and so on (Neander 2017; Shea 2018; Lee 2018). Moreover, the class of realistic theories of mental representation we are considering takes the content of such representations to be inherent — this is, not supplied by external attributions.

Second, it must be shown how this structure plays a causal role in cognition in virtue of its content.⁸ This is because, as Sprevak (2011) reminds us, if content “is just along for the ride, and does no causal or explanatory work, then there seems no reason to assume that the state in question is specifically representational, rather than, say, a causal relay with the same effects” (p. 670). As such, nothing will be a mental representation if it does not play a distinctive causal-explanatory role in cognition in virtue of its content. This is a pivotal point, for having content is what distinguishes mental representations from other states that can also causally contribute to bringing about cognitive activity.

It is claimed that S-representations can answer the job description challenge. Even Ramsey (2018), who originally issued the challenge, tells us that S-representations will form “an essential part” of our best representational-cum-computational accounts of cognition (p. 269, see also Opie and O’Brien 2015; Williams and Colling 2018; Lee 2018). Likewise, Piccinini (2018) observes that there is “an emerging consensus that the best way to understand representation in the context of cognitive explanation is structural” (p. 5).

What, apparently, makes S-representations fit for such special theoretical duty? Why do so many theorists think that S-representations manage to answer the job-description challenge while all other theories of mental representation fail? Allegedly, S-representations have unique properties, and, in understanding them, it becomes evident how such mental representations have inherent content and how those contents can play a genuinely causal role in driving cognition.

The favoured position in the literature is that S-representational vehicles “are contentful in virtue of resembling their represented objects” (O’Brien 2015a, p. 9). Defenders of S-representations claim that structural similarities “ground”

mental representational content (O'Brien and Opie 2004, p. 6, 8-14; O'Brien 2015a, p. 10). Following Von Eckardt (1993), this means that structural similarities are taken to be the “set of properties or relations that determine the semantic properties in question” (p. 199).

The rationale is as follows. The properties of a given S-representational vehicle, R , cause it to be structurally similar to some target state of affairs, T . Because R can mirror the structure of T more or less accurately, structural similarity entails accuracy conditions. Accuracy conditions are taken to entail content. Therefore, structural similarity is taken to entail content. Thus, S-representationalists conclude, the fact that R structurally mirrors T entails that R contentfully represents T .

According to this line of reasoning, the contentful properties of S-representations are supplied by the vehicular properties that make them structurally similar to their targets. If this is right, the contents of S-representations, unlike the contents of symbolic representations, are inherently and non-arbitrarily related to the properties of their vehicles (Williams 2017; Lee 2018).

Moreover, since it is assumed that such structural similarity is causally relevant for the success or failure of a given cognitive function, if structural similarity entails content then it follows that cognitive systems whose behavior is causally guided by the structural similarities holding between its S-representational vehicles and their targets are cognitive systems whose behavior is causally guided by mental representational contents. Representational contents fixed by structural similarities are, according to S-representationalists, unproblematically causally potent of cognition.⁹

With this in mind, O'Brien and Opie (2010) tell us that cognitive processes that involve S-representations:

are driven by the very properties that determine the content of its vehicles. In this sense, an [S-representational system] is not a mere semblance of a semantic engine—it's the real thing. *Any organism whose inner processes are analog in nature is causally indebted to the semantic properties of its inner states* (p. 127, emphasis added).

Thus, according to their defenders, S-representations are not only consistent with what we know about cognitive neuroscience, but they are better equipped to solve the classical challenges faced by other representational theories in cognitive science. In light of this, S-representations have been invoked in virtually every area of cognitive science, including classic symbolic computation (Gallistel 1990; Gallistel and King 2009), connectionism (O'Brien and Opie 2006, 2009), and predictive coding (Gładziejewski 2016; Kiefer and Hohwy 2017).

3. The Case Against Structural Representations

A standard, first-pass objection to S-representational theories of mental representation is that, even if we accept all that has been said so far, any content an S-representation might have is indeterminate.¹⁰

To see this, consider, again, cartographic maps — the paradigm example of non-mental S-representations. A map might be said to contentfully represent Sydney if the metrical relations among its constituent elements mirror the metrical relations between the relevant features of Sydney. Yet it can be the case that the very same map mirrors, perhaps to a different degree, the spatial layout of New York City as well. If that is the case, it would seem that S-

representationalists have to say that the same map represents both Sydney and New York City. Simply put, if the representational content of a map is wholly and uniquely determined by what it structurally mirrors, then what it represents, assuming it represents at all, is indeterminate.

Two main solutions to this content-specificity problem have been offered in the literature. The first solution proposes that the content of an S-representation is fixed, not only by what is being mirrored, but by whatever the S-representation targets when it is used to guide cognitive activity. Here we can assume, along with Godfrey-Smith (2006), that “[t]he target of a map is just whatever the map is in fact used to deal with” (p. 58).¹¹ The second solution, instead, proposes that S-representational contents get determined etiologically – that is, by what they were selected for dealing within the history of the cognitive agent’s engagements.¹²

We should not be distracted by these possibilities. A much stronger objection to S-representations focuses not on what makes the putative content of structural similarities determinate, but on whether structural similarities have any inherent content at all. Ultimately, the S-representationalist package is only tenable if it can account for the source of S-representational contents. In other words, what is required is a naturalistic account of the representational content of S-representations.

Recall that to be a structural representation, “a state must belong to a system of states that bear a second-order [structural] similarity to their targets … and the states must guide action based on their similarity to their target” (Piccinini 2018, p. 3). We are also told that “when a system’s internal states satisfy [the above conditions] they qualify as representations in a robust sense, which possess semantic content by the lights of a naturalistic theory of semantic content” (Piccinini 2018, p. 3).

Yet, the pivotal question for S-representationalists is whether the holding of structural similarity relations – on their own – suffices for one state of affairs to specify, refer or describe something about another state of affairs in a way that can be true or false, accurate or inaccurate, veridical or non-veridical. In other words, whether the fact that a particular item structurally mirrors another item suffices for the former to contentfully represent the latter.

On close examination, we contend, even though structural similarities might be said to ground content in the sense of enabling contentful evaluations and inferences, there is no reason to believe that such structural similarities, in-and-of-themselves, are inherently contentful.

Consider the following case. Against the backdrop of certain practices, we can use variations in the level of the tides in a particular region to make inferences and say something true or false, for example, about changes in the position of the Moon relative to the Earth. In any particular case, attending to the level of tides may or may not be an accurate or reliable way of keeping track of or saying something true about the position of the moon. The same goes for the way we use variations in the height of the liquid column of a mercury thermometer to make inferences about changes in the ambient temperature of the room. In both cases, certain structural correspondences or similarities must hold for the claims to be true and the inferences valid. This is so even though there is no reason to suppose that the structural correspondences in question must, themselves, represent things accurately or inaccurately.

In other words, it does not follow from the fact that we can make truth-evaluative claims based on structural similarities holding between two items, *A* and *B*, that *A* contentfully represents something that might be true or false about *B*. It is thus logically confused to suppose that the correspondences or structural similarities in question must be contentful in order for us to rely on

them in ways that make it possible to be right or wrong about things contentfully.

Following Gładziejewski and Miłkowski (2017), we can imagine a cognitive agent whose behavior is endogenously controlled by a mechanism that is sensitive to the fact that some internal states of the agent change concurrently with changes in the external environment. In such a scenario, we agree that a promising way to explain how the agent manages to cope with the environment “is to point to the [structural] similarity between its internal processes and the processes in the environment” (Gładziejewski and Miłkowski 2017, p. 351). We disagree, however, that this would be a case in which the behavior of the agent is causally guided by representational content. Again, this is because we have no reason to believe that structural similarities that meet the stated conditions suffice for mental representations with contentful properties. The mere fact that structural similarities can play a causal role in enabling successful acts of cognition does nothing to establish that structural similarities are contentful.

So far, we have been given no reason to suppose that structural similarities are sufficient for or entail content (see Goodman 1968, Fodor 1987, Hutto 2008 for similar claims). If structural similarities do not suffice for or entail content, then *a fortiori* they do not get their cognitive work done in virtue of possessing content.

In this light, if S-representationalists hope to meet the job description challenge they must do more than simply presuppose that neurally-based structural similarities are contentful; they must explain *how* and *why* structural similarities are contentful. To answer this, in essence, requires answering the general challenge posed by the Hard Problem of Content, or HPC, articulated by radical enactivists (Hutto and Myin 2013). What is needed is a substantial theory that accounts for S-representational content.

One option, at this point, is to attempt to explain S-representational content by appeal to some further theory of mental representational content. Most philosophers turn to teleosemantics to do this crucial work. For example, Thomson and Piccinini (2018) tell us that teleosemantics is “the best-developed and most plausible theory of representational content in biological systems” (p. 194).

Yet teleosemantics faces well-known shortcomings (Stich 1990; Fodor 1990; Rosenberg 2015; Burge 2010, Hutto and Myin 2013; Hutto and Satne 2015). To use a familiar example, whereas teleosemantics can explain why frogs snap their tongues in presence of black dots, it lacks the resources to specify, unequivocally, whether the frog’s visual system represents them as “black dots,” “flies,” “food,” and so on. Thus, even though teleosemantics provides the required resources to explain how organisms come to systematically target certain aspects of the world and not others, it fails to deliver a robust theory of representational content according to which entities are picked out intensionally or under a description.

We will not rehearse those arguments again here. Instead, we will assess Lee’s (2018) attempt to address those concerns afresh and head-on. Concurring with the above analysis, Lee (2018) holds that existing S-representational accounts have only “touched upon the issue of how to think about content” (p. 2). In an attempt to do better on this score, he aims to “show in detail how we can provide a naturalised understanding of content that dispels the strongest accusations of the antirepresentationalist” (p. 2).

Along with other defenders of S-representations, Lee (2018) holds that structural similarities are “what underwrites the representation’s degree of accuracy (its ‘accuracy value’)” (p. 2). But this, again, is insufficient to show that S-representations have content. Notably, Lee acknowledges that

correspondences of the sort structural similarities embody do not explain the source of representational content. He agrees that answering the HPC requires providing an account of S-representational content that does not rely solely on the notion of information-as-covariance.

At this juncture, Lee looks to the notion of non-natural information as a promising way to address the HPC since, arguably, this notion allows for the possibility of misrepresentation.¹³ He tells us that:

a non-natural information bearer is distinct from a natural information bearer in that it is both potentially decouplable from the conditions it bears information about, and the conditions it bears information about may be false. Yet both intuitively, and implicitly within the practice of cognitive science, non-natural information remains ‘informative’ (p. 9).

But, ultimately, Lee’s (2018) strategy falls short of providing a straight answer to the HPC. This is so because, as he admits, in these debates “there is no adequate theory which justifies the presence of non-natural information” (p. 10).

In the end, instead of answering the HPC, Lee (2018) argues that considerations about the explanatory role of non-natural information in cognitive science give us reason to question the legitimacy of what the HPC demands. As he holds, faced with a choice between recognizing the centrality of non-natural information to explanations of cognitive neuroscience or the need to answer the HPC, we should question “the severity of the HPC” (p. 10).

In sum, rather than explaining how S-representations can have content, Lee (2018) motivates acceptance that they do so by focusing on the explanatory work allegedly achieved by S-representational contents. As he argues, given

the kind of explanatory work earmarked for S-representations, we are warranted in assuming that S-representations have content.

Specifically, Lee (2018) holds that if S-representations are involved in error-detection work then we are justified, in light of explanatory need, in assuming that they are contentful. Why so? In his own words:

If one's theory of a system features an S-representation with a feedback component, whereby the system adjusts its behaviour based on a mismatch between an S-representation and some feature of a task, then this mismatch ... provides further justification for thinking that error, therefore accuracy conditions (therefore content), contributes to our understanding of how the mechanism works. (p. 12)

The important thing to notice is that not all mismatches entail representational errors. In fact, a key claim of radical enactivism (Hutto and Myin 2013, 2017) is that it is possible for some cognitive activities of agents to be "pragmatically mis-aligned, insensitive to certain features of the environment in a way that causes their efforts to fail" (Roelofs 2018, p. 246). Therefore, when explaining why such cognitive activities go wrong it would be a mistake to assume that cognitive systems must always go wrong by representing things wrongly. This is true even in those cases in which a pragmatic misalignment is brought about by a failure of a system's internal structures to mirror those of some target sufficiently well (see Kirchhoff and Robertson 2018).

Moreover, a system may be sensitive to such failures and may respond by effecting a stronger match of its internal states to some target, thereby improving its chances of successful outcomes. Yet a system can do all of this

without any part of it specifying or contentfully representing that certain conditions hold.

Imagine a set of keys, but only one of them can open a particular lock. Each key has its own unique geometry such that only one ‘fits’ the lock sufficiently well to open it. We can imagine a system designed to attempt to open the lock by using those keys. We can also imagine that the system is sensitive to the resistance of being unable to turn the key as a signal to try another key, and that it will continue with such a strategy – using one key after the another – until it succeeds.

What this simple example shows is that it is possible to be sensitive and responsive to mismatches in the completing of tasks without having to contentfully represent such mismatches (see Miłkowski 2015 for additional examples). This being the case, it is possible to explain a system’s sensitivity to mismatches and its capacity to make corrective adjustments to them without calling on the notion of representational content.

In conclusion, even if we accept that there is similarity-based cognition at the neural level, we argue that appealing to structural similarities provides no new resources for overcoming the hard problem of content – namely, that of accounting for the origins of content naturalistically – and, *a fortiori*, helping us to understand how such putative contents could possibly make a causal difference to cognition. At the same time, *pace* Lee (2018), we see no compelling reasons for believing that there is an indispensable explanatory need to posit the existence of such representational contents, even in the case of cognitive systems that are equipped with error-detection mechanisms.

4. A Further Case Against Structural Representations

Might contentless structural similarities still play a causal-explanatory, and yet properly representational, role in cognitive science even if they are contentless? Some proponents of S-representations seem to think so.

Gładziejewski and Miłkowski (2017) maintain that the new emphasis on the explanatory value of structural similarities in cognitive neuroscience affords “an opportunity to develop, strengthen, and indeed *reform the mainstream understanding of what representations are*” (2017, p. 338, emphasis added). For our purposes, the crucial adjustment would be to divorce the notion of mental representation from any and all connections with the notion of content.¹⁴ Others, such as Jacobson (2003, 2015), argue that no revision is needed. This is because, by her lights, the notion of content is a philosophical invention, and cognitive neuroscience has long been in the business of positing contentless representations.¹⁵

How can we understand S-representations as contentless but nevertheless representational? What justifies thinking of such structural similarities as representations if we assume that they lack content?

A familiar argument for this conclusion defends the idea that structural similarities function as maps or models in cognitive systems at the neural level. According to this view, since maps and models are primary examples of everyday non-mental representations, this gives us reason to regard neurally-based structural similarities as paradigm mental representations.

Ramsey (2016) articulates this view in the following terms:

If we think of mental representations not as indicators but instead as something more like elements of maps, models or simulations, then we can at least get the outlines of a story about how a part of the brain could actually function in a representational manner. (p. 7)

Yet for this argument to have bite neurally-based structural similarities must be more than merely map-like or model-like. They must actually function as maps or models in cognition. Why so? The reason is obvious. Everything is map-like or model-like to some degree. Hence, without further qualification, it is trivial to satisfy this criterion. Consider that humans are protozoa-like, and vice versa in that both humans and protozoa have cells.

So, the real question is: Are structural similarities actually *exploited* by a cognitive system or some part of it as models, maps, or surrogates of distal targets? Again, this question is important because, if they are, then there would be strong reasons to think, by analogy, that they are playing a *bona fide* representational role in cognition.

Reasons have been given for thinking that neurally-based structural similarities should be thought of play the role of a map in cognition. For example, elaborating on the place cells, Ramsey (2016) maintains that:

Insofar as these neural transformations implement a coordinate geometry during navigation that reflects the structure of the items and properties of the environment, *it is perfectly natural and, more importantly, explanatorily beneficial to regard such a system as functioning as a map*. Specific elements of the system are thus functioning as representations of features of the target domain (p. 8, emphases added).

At this juncture, however, we must be careful not to conflate evidence that the neural structures in question have map-like properties with evidence that brains or cognitive systems are using such structures as maps.¹⁶ What needs to be shown is that the way the rat's brain uses place cells in order to guide navigation is analogous, in a full-bodied sense, to the way agents use

cartographic maps. This is pivotal since, again, robustly satisfying the analogy is what, allegedly, secures that place cells are playing a representational role.

Some philosophers argue that new empirical findings regarding place cells motivate thinking that the analogy with maps holds strongly in this case (see Gładziejewski 2015; Gładziejewski and Miłkowski 2017, Shea 2018). These empirical findings show that, in some circumstances, hippocampal neurons fire in advance of action. As Gładziejewski and Miłkowski (2017) argue, this evidence supplies a compelling reason to think that place cells are not only responsive to the current location of the animal but, rather, that they are used by the rats' brains as surrogates to plan potential routes towards a target location.

For example, after having recorded neural activity in the CA3 region of the hippocampus of rats engaging in decision tasks, Johnson and Redish (2007) discovered that many sequences of spikes, or “sweeps,” fired by place cells during theta oscillation correspond to locations ahead of the rat.

For their experiment, they used two different mazes—one (called “multiple-T maze”) consisting of four T choice points and another (called “cued-choice maze”) consisting of a single T turn. Both mazes had two return trails with two places providing reward food, but only the feeders on one side were providing food each day. In addition, different cue sounds were played before the final turn (the so-called “choice point”), indicating which side would provide reward on each trial.

When rats reached the choice point, they faced a high-cost decision — taking the incorrect route means having to run approximately 3 m along the track before having another chance. Experimenters saw that, after being trained, rats often paused at the choice point, and that place cells corresponding to locations in both return trails fired while the individuals were standing still. Importantly,

the sweeps occurred separately — first for the cells corresponding to one side, and then the other — and they were not correlated to the orientation of the animal.

In a more recent experiment, Pfeiffer and Foster (2013) used a 2m x 2m open-field arena with 36 clearly demarcated locations. In the experimental task, rats had to alternatively forage for food rewards between randomly distributed locations and a stationary ‘home’ base. The location of the home base changed daily but remained constant within each day so that rats could remember it.

Importantly, Pfeiffer and Foster (2013) found that in the moments when a rat paused before taking a journey, sequences of activity in place cells not only sweep ahead of it, but they transiently predict the journey that the animal is about to take.¹⁷ Like in the previous experiment, future-oriented sweeps were seen to be independent of the rat’s orientation. As they explain, these discoveries “reveal a flexible, goal-directed mechanism for the manipulation of previously acquired memories, in which behavioral trajectories to a remembered goal are depicted in the brain immediately before movement” (Pfeiffer and Foster 2013, p. 78).

According to Gładziejewski and Miłkowski (2017), this body of empirical evidence reinforces the view that place cells are exploited in a way that is strongly analogous to the process of consulting a cartographic map. As they write: “the [hippocampal] map is internally manipulated [and] these manipulations are functional for the navigational mechanism in that they (presumably) serve as a basis for route planning” (p. 351). Likewise, Shea (2018) tells us that the evidence “suggests that rats use this prospective activity to plan the route they are about to follow” (p. 115).

Yet, again, a note of caution is needed. The experiments show that place cells sometimes activate in a future-oriented manner, and that such activity is

strongly biased toward the satisfaction of a goal. Yet even if the future-oriented activity of place cells plays a part in explaining how rats navigate to a location, it does not follow that the process in question equates to or involves route planning.

For this to be the case, forward sweeps would need to be involved in a process of evaluation so that they are used by other parts of the brain as surrogates of the available routes in order to choose the preferred one (Miłkowski 2015). This is something the experimenters themselves acknowledge:

Nonlocal forward [sweeps] are not sufficient for the consideration of future possibilities. Such consideration processes would also require mechanisms for the evaluation of nonlocal [sweeps] as well as mechanisms for flexible translation into behavior. (Johnson and Redish 2007, p. 12184)

The problem, however, is that it remains unclear whether, how and where this evaluation takes place. As Schmidt and Redish (2013) acknowledge, the hippocampus is thought to be part of a complex neural network that involves several brain structures, but it is unclear how the mentioned hippocampal activity interacts with the other brain structures in order to generate behavior. As they write, after these empirical findings “researchers must now explore what processes generate these place-cells sequences, and how they are used in recalculating the journey home” (p. 43; see also Pfeiffer and Foster 2013, p. 78).¹⁸

Therefore, however ‘natural’ it may be to gloss what is going on in the rat’s brain as a map-using process, considering the current available evidence, we should not assume that the rat’s brain engages in any kind of planning by means of surrogate reasoning in these cases. There is no evidence available that the neural activity that drives behavior in response to place cell firings is

anything like the process of consulting a cartographic map in order to navigate a location and plan a route. We agree with Godfrey-Smith (2006) that:

It is natural from the scientist's point of view to say that the rat is using X [the structure of place cells] as a guide to Y [the maze], but as far as the mechanics of the situation are concerned, the 'as guide to Y' claim seems extraneous. (p. 51; see also Rosenberg, 2018, p. 138)

Thus, without further evidence to show why we should think of these processes as robustly, and non-metaphorically, involving the use of a mental or cognitive map, we conclude that calling neurally-based structural similarities maps or models is unsupported. It follows that their representational status, in so far as it allegedly depends on they being used as models or maps, is unjustified.

5. A Radical Enactive Take on SBC

So far, we have argued that there is no reason to assume that the existence of causally potent neurally-based structural similarities entails the existence of mental representations. An important consequence of our analysis is that non-representational accounts of cognition can embrace neurally-based SBC. In what follows, we briefly explore how neurally-based SBC might be construed under the auspices of one radical view of cognition – namely, the radical enactive account of cognition, or REC, advanced by Hutto and Myin (2013, 2017).

A signature idea of REC is that cognition does not always and everywhere involve or entail representational content. When engaged in perceptual-motor tasks, for example, REC proposes that organisms can detect, track, and interact

with salient aspects of the environment by sensing and responding to the covariant information available in it, but they need not internalize or represent such information.

REC challenges the longstanding assumptions that the brain is either the seat of cognition or plays the chief role in enabling cognitive activity in virtue of neural states representing aspects of the environment. This, however, is not to deny that, at least for certain kinds of cognizers, neural activity plays a fundamental part in cognition, including basic forms of cognition. REC assumes that the primary function of dispositional patterns of neural excitation and inhibition is to coordinate the dynamically unfolding responses of organisms as they attune and adjust to environmental offerings in completing specific tasks (see also Engel et al. 2013; Gallagher 2017). We contend that adopting such a non-representational, action-oriented approach to neurodynamics is compatible with accepting that neurally-based structural similarities may play a pivotal role in explaining centrally important forms of intelligent and target-oriented behavior.

In understanding the kind of work the brain does in cognition REC draws on Anderson's theory of neural reuse (Anderson 2014). For Anderson, different parts of the central nervous system — at different scales, individual neurons, neural networks — are used and reused to accomplish different cognitive tasks. When this occurs, the various regions of the brain are temporarily soft-assembled into functional units or systems. Accordingly, brains causally contribute to enabling intelligent behavior in a variety of circumstances by “putting [the same neural structures] together in different patterns of functional cooperation” (p. 5).

Following Anderson's theory, Hutto et al. (2017) have proposed that brains are fundamentally “protean.” The Protean Brain Hypothesis conjectures that

brain structures are functionally malleable and context-dependent. Brains involving cognitive systems, according to this hypothesis, make use of neural structures in inventive, on-the-fly improvisations to meet the system's needs in specific circumstances.

Importantly, for our purposes, the way neurally-based structural similarities help to explain certain instances of intelligent behavior can be understood in conjunction with neural reuse and the Protean Brain Hypothesis. Accordingly, the dynamic activity of the central nervous system can play a part in enabling intelligent behavior by temporally reconfiguring already existing neural structures in order to resemble specific aspects of relevant targets. This is a particular way of understanding how the dynamics of the central nervous system can make a causal contribution to the intelligent behaviour of cognitive systems without assuming that the brain is in the business of representing the external world.

Crucially – focusing again on the parade case of place cells – the possibility that rat brains are using the forward-orientated firing of place cells for route planning is not the only available interpretation of the empirical evidence. Following Gallagher (2017), we contend that the fact that place cells fire in advance of action can be alternatively understood as “a constitutive part of the action itself, understood in diachronic, dynamical terms, rather than something decoupled from it” (Gallagher 2017, p. 14). On this view, anticipatory neural activity, operating on elementary timescales, can play a part in engendering larger-scale temporally extended cognitive activity (Stepp et al. 2011). Importantly, this can be the case even if the neural activity in question is not used by the rest of system as a separate process that fuels further distinct acts of reasoning or inference about possible courses of action.¹⁹

6. Conclusion

In this paper, we have challenged the popular suggestion that the existence of neurally-based SBC gives us reason to believe in S-representations. We offered two arguments against this view.

First, we argued that there is no reason to suppose that structural similarities, in themselves, suffice for or entail content. Hence, there is no reason to think that structural similarities do their cognitive work in virtue of possessing contents. If so, then S-representations fail to answer the job description challenge.

Second, we have also shown that there is no compelling evidence that neurally-based structural similarities function in a robust sense like maps when doing their cognitive work. Focusing on the parade case of place cells, we argued that the existing empirical evidence regarding the future-oriented activity of place cells does not provide compelling grounds for thinking that such neurally-based structural similarities are being used as maps, models, or surrogates of the external world.

Putting all of this together, we conclude that there is no reason to assume that the existence of causally potent neurally-based structural similarities entails the existence of mental representations. Therefore, *pace* Thomson and Piccinini (2018), we should not infer from the fact that modern techniques in experimental neuroscience allow us to observe structural similarities doing causal work in cognition that we are observing S-representations in action.

A crucial consequence of our analysis is that radical embodied, non-representational accounts of cognition can embrace neurally-based SBC. We have defended this view in the context of REC and the Protean Brain Hypothesis (Hutto and Myin, 2013, 2017; Hutto et al. 2017). Importantly, going

radical on this score enables us to acknowledge the value the core machinery neurally-based SBC while characterizing it in more deflationary terms. It should be clear, however, that in challenging the representational status of SBC, and in taking a non-representational, action-oriented approach to neurodynamics, REC is breaking with business-as-usual cognitive science.

References

- Anderson, A. (2014). *After phrenology. Neural reuse and the interactive brain*. Cambridge, Mass.: MIT Press.
- Bechtel, W. (2016). Investigating neural representations: The tale of place cells. *Synthese*, 193(5), 1287–1321.
- Burge, T. (2010). *The origins of objectivity*. Oxford: Oxford University Press.
- Cummins, R. (1994). Interpretational semantics. In S. P. Stitch & T. A. Warfield (Eds.), *Mental representation: A reader* (pp. 297–298). Cambridge, MA: Blackwell.
- Egan, F. (forthcoming). A deflationary account of mental representation. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* Oxford: Oxford University Press.
- Fodor, J. A. (1987). *Psychosemantics. The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge, Mass.: MIT Press.
- Gallagher, S. (2017). *Enactivist interventions. Rethinking the mind*. Oxford: Oxford University Press.
- Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: why cognitive science will transform neuroscience*. Malden, MA: Wiley-Blackwell.
- Garzón, F. C., & Rodríguez, Á. G. (2009). Where is Cognitive Science Heading? *Minds and Machines*, 19(3), 301–318.

- Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, 193(2), 559–582.
- Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: causally relevant and different from detectors. *Biology & Philosophy*, 32(3), 337–355.
- Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In D. Papineau & G. MacDonald (Eds.), *Teleosemantics: New philosophical essays* (pp. 42–68). Oxford: Oxford University Press.
- Godfrey-Smith, P. (2009). Representationalism reconsidered. In D. Murphy & M. A. Bishop (Eds.), *Stich and His Critics* (pp. 30–46). Malden, MA: Wiley-Blackwell.
- Godfrey-Smith, P. (2012). Signals, Icons, and Beliefs. In D. Ryder, J. Kingsbury, & K. Williford (Eds.), *Millikan and Her Critics* (pp. 41–62). Malden, MA: Wiley-Blackwell.
- Goodman, N. (1968). *Languages of art*. London, UK: Oxford University Press.
- Hutto, D. D. (2008). *Folk Psychological Narratives. The sociocultural basis of understanding reasons*. Cambridge, MA: MIT Press.
- Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism : Basic minds without content*. Cambridge, MA: MIT Press.
- Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge, MA: MIT Press.
- Hutto, D. D., & Myin, E. (2018). Much ado about nothing? Why going non-semantic is not merely semantics. *Philosophical Explorations*, 21(2), 187–203.
- Hutto, D. D., & Myin, E. (forthcoming). Deflating deflationism about mental representation. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* Oxford: Oxford University Press.
- Hutto, D. D., Peeters, A., & Segundo-Ortin, M. (2017). Cognitive ontology in flux: the possibility of protean brains. *Philosophical Explorations*, 20(2), 209–223.
<https://doi.org/10.1080/13869795.2017.1312502>
- Hutto, D. D., & Satne, G. (2015). The natural origins of content. *Philosophia*, 43(3), 521–536.

- Jacobson, A. J. (2003). Mental representations: What philosophy leaves out and neuroscience puts in. *Philosophical Psychology*, 16(2), 189-203.
- Jacobson, A. J. (2015). Three concerns about the origins of content. *Philosophia*, 43(4), 625-638.
- Johnson, A. & Redish, A. D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *The Journal of Neuroscience*, 27(45), 12176-12189.
- Kirchhoff, M. D., & Robertson, I. (2018). Enactivism and predictive processing: a non-representational view. *Philosophical Explorations*, 21(2), 264-281.
- Knierim, J. J. (2015). From the GPS to HM: Place cells, grid cells, and memory. *Hippocampus*, 25, 719-725.
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412.
- Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, Mass.: MIT Press.
- Miłkowski, M. (2015). Satisfaction conditions in anticipatory mechanisms. *Biology and Philosophy*, 30, 709-728.
- Morgan, A. (2014). Representations gone mental. *Synthese*, 191(2), 213–244.
- Neander, K. (2017). *A mark of the mental: In defense of informational teleosemantics*. Cambridge, MA: MIT Press.
- Lee, J. (2018). Structural representations and the two problems of content. *Mind & Language*, 1-21.
- O'Brien, G. (2015a). How does mind matter? Solving the content causation problem. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. <https://doi.org/10.15502/9783958570146>
- O'Brien, G. (2015b). Rehabilitating resemblance redux. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. <https://doi.org/10.15502/9783958571136>

- 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* (pp. 1–20). Oxford: Elsevier.
- O'Brien, G., & Opie, J. (2006). How do connectionist networks compute? *Cognitive Processing*, 7(1), 30–41.
- O'Brien, G., & Opie, J. (2009). The role of representation in computation. *Cognitive Processing*, 10(1), 53–62.
- O'Brien, G. & Opie, J. (2010). Representation in analog computation. In A. Newen, A. Bartels, & E. M. Jung (Eds.), *Knowledge and representation* (pp. 109-129). Stanford: CSLI Publications.
- O'Brien, G., & Opie, J. (2015). Intentionality lite or analog content? *Philosophia*, 43(3), 723–729.
- O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1), 78–109.
- O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1), 171–175.
- O'Keefe, John, & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.
- Piccinini, G. (2018). Computation and representation in cognitive neuroscience. *Minds and Machines*, 28(1), 1-6.
- Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cells sequences depict future paths to remembered goals. *Nature*, 497(7447), 74-79.
- Ramsey, W. M. (2016). Untangling two questions about mental representation. *New Ideas in Psychology*, 40, 3–12.
- Ramsey, W. M. (2018). Maps, models and computational simulations in the mind. In M. Sprevak & M. Columbo (Eds.), *Handbook of the Computational Mind*. Rutledge (pp. 259-271).

- Ramsey, W. M. (forthcoming). Defending representational realism. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* Oxford: Oxford University Press.
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge, UK: Cambridge University Press.
- Rescorla, M. (2016). Bayesian sensorimotor psychology. *Mind and Language*, 31(1), 3-36.
- Roelofs, L. (2018). Why imagining requires content: A reply to a reply to an objection to radical enactive cognition. *Thought: A Journal of Philosophy* 7(4), 246-254.
- Rosenberg, A. (2015). The genealogy of content or the future of an illusion. *Philosophia*, 43(3), 537-547.
- Rosenberg, A. (2018). *How history gets things wrong: The neuroscience of our addiction to stories*. Cambridge, Mass.: MIT Press.
- Sachs, C. B. (2018). In defense of picturing. Sellars's philosophy of mind and cognitive neuroscience. *Phenomenology and the Cognitive Sciences*.
- Schmidt, B., & Redish, A. D. (2013) Navigation with a cognitive map. *Nature*, 497(7447), 42-43.
- Shagrir, O. (2012). Structural representations and the brain. *The British Journal for the Philosophy of Science*, 63(3), 519–545.
- Shea, N. (2013). Millikan's Isomorphism Requirement. In D. Ryder, J. Kingsbury, & K. Williford (Eds.), *Millikan and Her Critics*. Malden, Mass.: Wiley-Blackwell.
- Shea, N. (2014). Exploitable isomorphism and structural representation. *Proceedings of the Aristotelian Society*, 114, 123-144.
- Shea, N. (2018). *Representation in cognitive science*. Oxford: Oxford University Press.
- Sprevak, M. (2011). Review of William M. Ramsey. Representation reconsidered. *The British Journal for the Philosophy of Science*, 62(3), 669–675.
- Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, 3, 425–437

- Swoyer, C. (1991). Structural representation and surrogate reasoning. *Synthese*, 87. 449-508.
- Thomson, E., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines*, 28(1), 191-235.
- Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, 55(4), 189–208.
- Von Eckardt, B. (1993). *What is cognitive science?* Cambridge, MA: MIT Press.
- Williams, D. (2017). Predictive Processing and the Representation Wars. *Minds and Machines*. <https://doi.org/10.1007/s11023-017-9441-6>
- Williams, D., & Colling, L. (2017). From symbols to icons: the return of resemblance in the cognitive neuroscience revolution. *Synthese*. <https://doi.org/10.1007/s11229-017-1578-6>

Acknowledgements: We would like to express our gratitude to the participants at the Neural Mechanisms Online Webinar Series (April 5th, 2019) for their useful comments and suggestions on an earlier version of this paper. Research for this article was supported by the Australian Research Council Discovery Project “Mind in Skilled Performance” (DP170102987).

¹ Swoyer (1991) illustrates this familiar phenomenon with the following example: “By examining the behavior of a scale model of an aircraft in a wind tunnel, we can draw conclusions about a newly designed wing’s response to wind shear, rather than trying it out on a Boeing 747 over Denver. By using numbers to represent the lengths of physical objects, we can represent facts about the objects numerically, perform calculations of various sorts, then translate the results back into a conclusion about the original objects. In such cases we use one sort of thing as a surrogate in our thinking about another, and so I shall call this *surrogate reasoning*” (p. 449, emphasis original).

-
- ² It is common in the literature to depict the mapping or mirroring relations in terms of isomorphism. However, current examples in neurocomputational theories of cognition appeal to highly abstract structure-preserving mapping relations that are considerably weaker than isomorphism (see Neander 2017, p. 176; see also Gładziejewski 2016; Morgan 2014). For our purposes, we can remain neutral with respect to this discussion. We will speak more generally and inclusively of structural similarities or resemblances (see O'Brien and Opie (2004) and Shea (2013, 2018) for a technical and detailed analysis of these notions).
- ³ It is not obvious that place cells constitute a Cartesian coordinate system. For example, Bechtel (2016) has argued that “[w]hereas in a cartographic map the spatial locations between representations correspond, albeit only approximately and with distortions, to the spatial relations between the places represented, this is not true of the map realized in place cells” (p. 1297, emphasis added). Shea (2014) raises similar doubts, observing that “[t]he mechanism depends on place cell firing correlating reliably with location, but not on any relation between different place cells, nor on spatial relation between locations” (p. 126, emphasis added).
- ⁴ “Explanations that invoke S-representations should thus be construed as causal explanations that feature facts regarding similarity as an explanans and success or failure as an explanandum. To exploit structural similarity in this sense is to use a strategy whose success is causally dependent on structural similarity between the representational vehicle and what is represented” (Gładziejewski and Miłkowski 2017, p. 340, emphasis added).
- ⁵ The relationship between similarity and success is not a straightforward one. Consider a cartographic map. A cartographic map does not fully replicate the terrain it is meant to represent. On the contrary, it simplifies it – only including elements that are relevant for the function it was designed to achieve. A map that resembles its target too much would become excessively complex and thus useless. The same rationale applies to S-representations. As Gładziejewski and Miłkowski (2017) note, “too much similarity can render the S-representation inefficient at serving its purpose” (p. 344).
- ⁶ In order to recognize the scope of the job description challenge it is important to mention that it does not just trouble S-representational theories in cognitive neuroscience. Instead, serious worries have been raised in its wake about the tenability of classical cognitivist's conjecture that cognition is rooted in digital computation. For, even if cognition proves to be digitally instantiated, there are deeper unanswered puzzles about how representational contents could be causally efficacious, rather than being systematically screened off from playing any causal explanatory role.

-
- ⁷ Accordingly, something only counts as a mental representation if it is “used to represent a … target as being a certain way that it might or might not be” (Neander 2017, p. 35; Rescorla 2016).
- ⁸ As Thomson and Piccinini (2018) present it, the received view is that “[f]or something to count as a representation, it must have a semantic content (e.g., ‘there is yogurt in the fridge’) and an appropriate functional role (e.g., to guide behavior with respect to the yogurt in the fridge)” (p. 193).
- ⁹ Invoking the much-discussed example of the thermostat, O’Brien seeks to demonstrate “the causal efficacy of content fixed by resemblance” (2015a, p. 9). As he tells us, the thermostat’s functioning is causally driven by the structural similarity holding between the curvature of the bi-metallic strip and the temperature of the room. Thus, if it is assumed that structural similarities are intrinsically contentful it would follow that representational contents can be causally efficacious of behavior.
- ¹⁰ Another, related, objection has to do with the fact that structural similarities, unlike representations, are symmetrical. A map structurally mirrors the layout of a city as much as the city structurally mirrors the layout of the map. If that is the case, S-representationalists have to conclude that the city represents the map too. To solve this problem, a number of authors have suggested to rethink the representation relation as a triadic relation, this is, as a relation that involves not only the representational state and its target, but also a representational user or consumer (Millikan 1984; O’Brien 2015a). With this condition at hand, we can now say that what makes the map a representation of the city, and not the other way around, is the fact that the map is being used or consumed as such by a cognitive agent or system.
- ¹¹ O’Brien (2015a, 2015b; see also O’Brien and Opie 2015) proposes a similar solution to the content-specificity problem, putting emphasis on the interpretive activity of users. According to this idea, an S-representational state R of a system S is a representation of T if S ’s responses to T are causally mediated by R . As he writes, “the behavioural dispositions of the system restrict the represented domain to [T], and the second-order resemblance relations determine what [features of T] each vehicle represents” (O’Brien 2015a, p. 11).
- ¹² Ramsey (2016) holds that a neural state R is a representation of T if T caused R to come about and acquire the structure it has. Thus, if a particular S-representation “was developed in an effort to learn how to navigate a specific maze, then it is that particular maze that is the target [of this S-representation]” (p. 7). Accordingly, in such cases, S-representational content is not fixed by structural similarity relations

solely, it is also fixed by the relevant causal relations that brought the S-representational vehicle into existence.

- ¹³ As Lee (2018) explains, “*x* bears natural information about *y*, iff *x* reliably covaries with *y*. In this case, *x*’s bearing information about *y* is dependent on a direct physical relationship. By contrast, *x* bears non-natural information about *y* iff *x* stands-in for *y*, where *x*’s tokening does not entail the truth of *y*. In this instance, *x*’s bearing information about *y* is not dependent on any direct physical relationship” (p. 8).
- ¹⁴ There is a tendency in the current literature to attempt to deflate the mainstream notion of mental representation. Egan (forthcoming) has suggested that we can treat representational content as an explanatory gloss. She proposes this maneuver as a way of retaining the notion of mental representation in the cognitive sciences while avoiding the seemingly intractable problem of providing a naturalistic explanation for the origin of representational contents. For detailed discussions of this kind of deflationary move see Ramsey (forthcoming) and Hutto and Myin (2018).
- ¹⁵ Interestingly, Jacobson justifies this idea by directly appealing to the explanatory role of similarity in cognitive neuroscience. As she writes: “With the rise of representational similarity and their elaboration of what representation in neuroscience amounts to, there seems no doubt now that cognitive neuroscientists have in mind a very different notion of representation … cognitive neuroscience is not employing contentful representations” (2015, p. 3).
- ¹⁶ As Shea (2018) explains, “the remarkable discovery of the location-specific sensitivity of place cells does not, by itself, show that rats have a cognitive map” (p. 115).
- ¹⁷ For their experiment, Pfeiffer and Foster (2013) recorded the activity of 250 place cells at short time scales (circa 20 ms). The sequences or sweeps measured by Pfeiffer and Foster occur during sharp-wave-ripple (SWR) events—this is, irregular burst of brief (100-200 ms) high-frequency (140-200 Hz) neuronal activity. Place cell sweeps during SWR events are traditionally associated to processes of memory consolidation during sleep.
- ¹⁸ Another, related, issue has to do with the relation between the discoveries of Pfeiffer and Foster (2013) and the ones of Johnson and Redish (2007). Schmidt and Redish (2013) ask: “what is the relation between these two planning phenomena? Does one negate the need for the other?” (p. 43)
- ¹⁹ There is a growing literature in cognitive neuroscience that holds that a non-representational reading of forward-oriented neural activity is feasible. According to these views, it is possible to understand the contribution of the future-oriented

neural activity to the system's behaviour without assuming that this neural activity represents future events (see, e.g., Kirchhoff and Robertson 2018; Gallagher 2017; Stepp et al. 2011).