The many problems with S-representation (and how to solve them)

Jonny Lee[a] [iD] (leejonathan.cw@gmail.com)
Daniel Calder[b] (DAC@christs-hospital.org.uk)

Abstract

The structural representation (S-representation) account provides an increasingly popular way of understanding the role and value of representation in cognitive science. Yet critics remain unconvinced that the account has the resources to rescue representationalism. This paper reviews problems faced by the S-representation account. In doing so, it offers a novel taxonomy that divides objections into two broad camps that ought to be disambiguated: 'conceptual' and 'empirical'. It further shows how these objections can be met, bolstering existing responses in the literature with novel solutions, thus strengthening the S-representation account. Finally, this paper suggests that the types of objections identified provide a generalisable taxonomy for a better understanding of any version of anti-representationalism. Thus, this review is of value to the representation debate more broadly.

1 Introduction

Contemporary debates over the role and value of representation in cognitive science revolve around the notion of 'structural representation', or 'S-representation' for short. The S-representation account characterises cognitive representation as a class of state, structure or mechanism component that guides the behaviour of a cognitive system by structurally resembling features of its task environment, thus playing a map- or model-like role.[1] Proponents argue this approach iden-

---

[a]University of Murcia.

[b]Christ's Hospital.

[1]For simplicity, and in accordance with literature from which we draw (e.g., Gładziejewski, 2015) we will refer to 'cognitive mechanisms' and 'representational mechanisms'. However, we do not assume that proponents of the S-representation account are committed to mechanistic explanations of cognition (e.g., Machamer et al., 2000), or that S-representations must be cast as mechanisms in the technical sense (e.g., see Cummins, 1996). Equally, if a mechanistic lens is the best way to view cognitive representation then our defense of S-representation only strengthens this perspective.

tifies a robust notion of representation that is relevant to our best explanations of cognition. Despite this promise, there remain several pressing criticisms of the S-representation account. This paper reviews possible objections and organises them within a novel taxonomy that is likely applicable to other concepts of representation. We offer responses to each objection, drawing on the existing literature as well as neglected considerations. In doing so, we both strengthen the S-representation account and illuminate the conceptual landscape surrounding the representation debate more generally.

Though anti-representationalism is sometimes presented as a monolithic position, we contend that there are at least five independent (albeit overlapping) types of objections to the S-representation account. We label these the 'a priori', 'function', 'content', 'best theory', and 'interpretation' objections. These fall under two broad categories, which we label 'conceptual' (a priori, function and content objections) and 'empirical' (best theory and interpretation objections). This taxonomy is both descriptive and prescriptive; it captures a variety of objections advanced by critics as well as distinctions between classes of argument that ought to be disambiguated (we assume the taxonomy is comprehensive but not that it is exhaustive). Our central contention is that the S-representation account has the resources to respond to each of these types of objections. We also take the taxonomy to be generalisable to any form of representationalism, beyond the S-representation account. Hence, the discussion is pertinent to all debates over the role and value of cognitive representation and will prove useful for pro- and anti-representationalists alike.

The paper proceeds as follows. §2 introduces the S-representation account. We pay special attention to clarifying the commitments of the S-representation account to swerve any potential strawmen. §3 sets out criticisms of the S-representation account falling under the 'conceptual' bracket—the 'a priori', 'function' and 'content' objections—and offers a response to each. §4 turns to 'empirical' concerns and how to counter them—covering the 'best theory' and 'interpretation' objections. We conclude by suggesting that our taxonomy is broadly applicable and caution those within the debate against conflating different types of arguments for anti-representationalism.

2   The S-representation account

Cognitive scientists ascribe internal representations to parts of cognitive systems. On closer inspection, however, representations are attributed according to different standards across cognitive science (for an overview of some of these inconsistencies, see Ramsey, 2007). Some have also questioned whether at least some so-called cognitive representations have a recognisably representational role (e.g., Facchin, 2021a), whether the concept of cognitive representation is coherent (e.g., Bennett & Hacker, 2007), and if we might be better off eliminating representation from cognitive science altogether (e.g., Hutto & Myin, 2012). The S-representation account offers a response to both the historic ambiguity of representation in cog-

nitive science, as well as sustained objections to some or all representation-based theories of cognition. The goal of the S-representation account is to articulate at least one set of jointly sufficient conditions for a theoretical posit or construct to count as a genuine representation, and in doing so, at least partially reflect representation's role and value in cognitive science. As such, the account need not exclude the possibility of other ways of representing; the criteria for S-representation are not necessary conditions for cognitive representation.

According to the S-representation account, representations are a class of mechanism component—at first pass, neural or computational structures within an information processing system—that play a map- or model-like role for a cognitive system. When a mechanism contains a component that determines the behaviour of a system (the 'user' or 'consumer') via structural correspondence with a task-relevant item (object, state, process etc.), the thought goes, then it functions like a map or model, i.e., a type of representation. For instance, as a mountaineer might use a cartographic map to navigate a mountain range by exploiting the structural similarities between the artefact and the geographical region, rats appear to use 'cognitive maps' to navigate their local environment by exploiting structural similarities between parts of their navigational system (primarily located in the hippocampus) and their local environment (O'Keefe, 1976; O'Keefe & Dostrovsky, 1971; Tolman, 1948). The S-representation account thus appeals to an analogy between a class of cognitive mechanism and a class of representational artefact. In Ramsey's (2007) terms, these mechanisms pass the 'job description challenge' (JDC)—they play a recognisably representational role—by functioning in a manner that resembles ordinary maps, models and the like. Thus, an "if it looks like a duck and quacks like a duck" principle of classification underlies the description of certain mechanisms as representations. This is akin to categorising hearts, certain mechanoreceptor cells and parts of the oceanic carbon cycle as 'pumps' because, like those ordinary artefacts from which we derive the term, they function to move fluids by mechanical action.

The S-representation account is elucidated through a set of conditions.[2] We follow others (particularly Głądziejewski, 2015) in identifying four key criteria for S-representation: structural correspondence, action guidance, decouplability and system-detectable error. As these have been discussed at length elsewhere, we restrict ourselves to a brief overview of each (Głądziejewski, 2015; Głądziejewski & Miłkowski, 2017; Lee, 2018; e.g., see O'Brien & Opie, 2004; Piccinini, 2020; Shea, 2018; Williams, 2017).

The central characteristic of an S-representation is its *structural correspondence* with elements in a system's task environment (broadly understood as the spatio-temporal region containing all variables bearing on the cognitive capacity being executed). As O'Brien & Opie (2004) summarise, "one system *structurally resembles*

---

[2]We take these conditions to be necessary and jointly sufficient for S-representation and, in turn, jointly sufficient for cognitive representation. Again, other conditions may suffice for representation.

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" (2004 summarise, 15). For instance, in the case of cognitive maps, cells fire selectively in response to particular locations, whilst the firing rate and strength of connections between cells correspond proportionally to the distances between features in the system's environment. Thus, it is assumed that structural resemblance between the cognitive mechanism and the environment is functionally relevant to the rat's success in navigating. Like ordinary maps, correspondence needn't be absolute for a mechanism to function as an S-representation; the structural correspondence between an ordinary artefact or cognitive mechanism and its target need only be as strong as required to complete the task.

Structural correspondence is thus only relevant in relation to *action guidance*. An S-representation corresponds to some task-relevant item and that correspondence is exploited by the system—as when rats exploit the correspondence between firing rates of place cells and features of an environment to aid navigation. Furthermore, in keeping with paradigmatic cases of representation, including maps and models, S-representations are potentially *decouplable* from that which they correspond to in guiding action—they must be usable 'offline'—as when rats route plan by exploiting cognitive maps. Finally, it should be possible for a system to *detect error* or a mismatch between the S-representation and the item it must correspond to for successful behaviour to occur—for example, an insufficient correspondence between a cognitive map and a rat's environment. Components then 'update' to instantiate a stronger correspondence and so raise the probability of success in subsequent tasks. Strictly speaking, the capacity to detect error and the capacity to update following error are conceptually independent and so might be understood as two conditions. In practice, however, paradigmatic S-representations, such as cognitive maps, are subject to amendment following the outcome of some behaviour and it is difficult to imagine how one might observe error detection without observing alteration to a mechanism in response to task failure. Hence, we treat these conditions as bundled together.

Having outlined the positive story, we can now highlight connections between the S-representation account and other pertinent ideas in the philosophy of cognitive science. Given our concern for objections to the account, we wish to avoid any strawmen lurking in the shadows. Thus, we offer three clarifications regarding the S-representation account. We believe more superficial criticism may be avoided by recognising these, and in the process, the concept of S-representation will become clearer.

The first clarification is that, though driven by an examination of scientific practice, the S-representation account is not an empirical theory. Rather, it is a generic account of what it means to be a type of cognitive representation. In turn, this can be used to understand the role and value of representation in cognitive science. For example, in assessing which theories in cognitive science posit anything resembling genuine representations, Ramsey (2007) concludes that the 'classical com-

putational theory of cognition'—roughly synonymous with cognitivism—is representational because it implicates a computational architecture that functions to model the world, in a sense analogous to more familiar models. By contrast, Ramsey concludes that connectionist theories of cognition employ no such model-like posits, nor anything else sufficiently representation-like (for pushback, see Shagrir, 2012). More recently, theorists have offered interpretations of the representational commitments of 'predictive processing' (PP) in terms of S-representation (e.g., Gładziejewski, 2016; for a partial review see Sims & Pezzulo, 2021, see §4.2 for discussion). The S-representation account is thus not an empirical theory but a set of conditions for a type of cognitive representation *qua* a type of theoretical construct or posit.[3]

Secondly, the account is neutral about the range of cognitive phenomena to be explained representationally. Sometimes, representations have been taken to be necessary for cognition (e.g., Adams & Aizawa, 2001). This is what Ramsey (2016) calls the 'representation demarcation thesis', that is, "the view that cognitive processes necessarily involve inner representations and cognitive theories must thereby be about the representational states and processes" (2016, p. 4). The S-representation account makes no such assumptions (cf. Cummins, 1996; Ramsey, 2007, 2016). Whether cognition is mostly representational, or whether it relies also on non-representational resources within a more mixed cognitive economy, depends on the details of our best theories and our broader conception of cognition. Furthermore, the possibility that some cognition within some systems involves S-representations does not preclude the existence of other cognitive systems which are wholly non-representational, nor the existence of systems which represent some other way. For instance, it may be that offline cognitive processes—so-called 'representation hungry' problems (Clark & Toribio, 1994)—that characterise 'higher cognition', such as reasoning about counterfactuals, do involve S-representations whilst online cognitive processes that comprise the minimal cognition of, say, bacterium do not (e.g., Duijn et al., 2006).

The third clarification concerns the compatibility between the S-representation account and some factions within 'embodied, embedded, enactive and extended cognition' ('4E cognition' or simply '4E'). Reviewing the long and complex relationship between representationalism and 4E is beyond the scope of this paper, but we should recognise that some theories and frameworks within 4E revolted against orthodox representationalism. However, the S-representation account is not orthodox representationalism. For one thing, as we saw above, it's not an empirical theory but an articulation of the generic properties belonging to a type of theoretical posit or construct. For another, as Gładziejewski & Miłkowski (2017)

---

[3]The metaphysical status of S-representations has been contested in recent years. We take the default view to assume that S-representations are a type of theoretical construct or posit but other interpretations claim they are observable entities (in particular, see Thomson & Piccinini, 2018). A review of possible positions on the metaphysics of cognitive representation must wait for another day.

point out, the notion of S-representation provides us with "an opportunity to develop, strengthen, and indeed reform the mainstream understanding of what representations are" (2017, p. 338). Such reform is driven, in part, by advances in cognitive science. As Williams & Colling (2018) note, the "cognitive neuroscience revolution", as outlined by Boone & Piccinini (2015), hails "a dramatic shift away from thinking of cognitive representations as arbitrary symbols towards thinking of them as icons that replicate structural characteristics of their targets" (2015, p. 1942). At the same time, many of the criticisms of representation found in 4E aim only at a particular understanding of cognitive representation—one that takes it to be discrete, language-like, brain-bound and purely descriptive (e.g., Clark, 1997, 2015b, 2015a), in other words, the very notion seemingly discarded by contemporary cognitive neuroscience.

We noted already that admitting some processes involve S-representations does not preclude the existence of non-representational processes, of the sort some 4E often concerns itself with.[4] Furthermore, the notion of S-representation is compatible with many advances within the representation debate which depart from traditional language-based formats and transformations. Taking just one example, the S-representation account seems congruent with Barsalou (1999) perceptual symbols theory which rejects the notion of amodal symbolic representations bearing arbitrary relations to their referents. Perceptual symbols theory stresses the dependence of representing structures underlying more 'cognitive' tasks, such as imagination, on the re-activation of sensorimotor processes giving rise to a kind of perceptual simulation. In turn, what a system can represent is constrained by the perceptual modalities embodied by the system. We cannot here offer a complete exploration of how S-representations are productively nested within a 4E-friendly research program [but for some initial attempts at how S-representation supports and vindicates aspects of 4E, (see Piccinini, 2022; Williams, 2018). For now, without downplaying the genuine tension between cognitive representation and certain quarters of 4E (e.g., see Varela et al., 1991), we contend that one's predilection for 4E approaches does not exclude the possibility of S-representations playing a constructive role within cognitive science (for sample pushback, see Hutto, 2013).

In closing our introduction to the S-representation account, it is worth acknowledging the possibility of 'explanatory pluralism' in cognitive science. According to pluralists, cognition can be understood through a variety of frameworks (e.g., Dale et al., 2009). Within such an explanatory melting pot, it may be that only some frameworks appeal to S-representations. For example, one may entertain an explanatory pluralism that legitimises both mechanistic and dynamical systems explanations, where the former posit model-like entities deserving of the name

---

[4]One recent example of this includes the dual role of 'model-free' and 'model-based' modes of decision making within PP, which suggests to us a partial but vital role for S-representations within an action-oriented framework (Clark, 2016, Chapter 8).

'representation' but the latter do not (see §4.1 for related discussion).[5] Notice that a version of pluralism in which representation plays a role only in some frameworks is not the same as instrumentalism about representation; pluralists may be realists about representation, denying only that such really existing entities are relevant to all types of explanation. Whether such explanatory pluralism is warranted depends on wider issues within philosophy of cognitive science. However, the possibility that representation can be quarantined to certain explanatory frameworks without infecting all cognitive science may relieve some sceptics of representation's explanatory power from the pressures of a zero-sum game.

3   Conceptual objections

There are at least five types of objections that arise from discussions of the S-representation account: the a priori, function, content, best theory and interpretation objections. This taxonomy reflects significant differences in argumentative strategy. There is overlap between these types, as we shall see, and anti-representationalists needn't think their position can be divided neatly into the categories we individuate—for instance, because the different types of objections reinforce each other. Nevertheless, we should avoid conflating distinctive sorts of arguments. In this section, we examine the a priori, function and content objections. These are chiefly 'conceptual' to the extent they question whether the criteria for so-called S-representation is sufficient for representation-hood. In §4, we turn to the best theory and interpretation objections. These are chiefly 'empirical' to the extent they question the relevance of S-representations for our best theories in practising cognitive science (and are thus sensitive to the emergence of new evidence), not what it means to be a representation.[6]

3.1   A priori objections

A priori objections fixate on the very notion of a subpersonal representation. For some critics, 'subpersonal representation' is a conceptual confusion; how could something like an electrical circuit or neural state literally represent anything? Notice that our first type of objection is a general one: it targets *all* accounts of representation. Since this 'a priori eliminativism' is concerned with the possibility of providing an account of cognitive representation it affects the S-representation account as a consequence of targeting every account of representation. As we shall see, there are broad counters available to any representationalist. However, there are also some specific considerations afforded by the S-representation account. In any case, it is important to consider a priori objections because if the S-

---

[5]Not everybody agrees that mechanistic and dynamical explanations are different forms of explanation (e.g., Brigandt, 2013). We return to this point in §4.1.

[6]For similar divisions, see Chemero (2009); Steiner (2014).

representation account is to be successful it must possess the capacity to respond. They are also important to acknowledge insofar as they help differentiate this class of conceptual concern (about representation generally) from other types of objections (about the S-representation account).

A priori eliminativism has its roots in a worry from ordinary language philosophy stemming from the assumption that representation implies some sort of user or consumer. In turn, the thought went, internal representations imply internal homunculi—mini agents who are required to interpret these representations (cf. Ryle, 1949; Wittgenstein, 1953). Hence, attributing psychological predicates to the brain was seen as a kind of category error.

Perhaps the clearest contemporary expression of this sentiment is found in Bennett & Hacker's (2007) attempt to exorcise 'personal level predicates' from the subpersonal level. Ascribing representations to the brain is, according to Bennett & Hacker's view, part of a larger practice of ascribing psychological attributes or predicates to the brain, such as when we talk about the brain constructing hypotheses, estimating probabilities or presenting arguments, as well as hearing, seeing and falling asleep (e.g., Bennett & Hacker, 2007, pp. 17, 21). Moreover, whether such practices are legitimate is principally a conceptual, philosophical question, not a scientific one. Their view is that ascribing psychological predicates to the brain is 'senseless'. The brain, in their words, is "*not a logically appropriate subject for psychological predicates*" (Bennett & Hacker, 2007, p. 21). Only humans and other agents are the proper domain of such attributes—only agents estimate probabilities or fall asleep. As Bennett & Hacker put it: "Psychological predicates are predicates that apply essentially to the whole living animals, not to its parts" (2007, p. 22). Hence, scientists and philosophers commit a 'mereological fallacy' when ascribing psychological predicates to parts of persons (Bennett & Hacker, 2007, p. 22).

Everyday life provides examples of constraints on literal reference; the wind does not literally whisper nor do fires rage—these are metaphors—and it is hard to imagine how these characteristically human behaviours *could* function as literal descriptions of the weather. Similarly, it is difficult to imagine neurons mourning, celebrating or conniving in any literal sense. The force of views like Bennett & Hacker's stems from generalising from such common sense cases of constraints on domains of reference to the conclusion that psychological predicates apply 'essentially' (Bennett & Hacker, 2007, p. 22) and 'paradigmatically' (Bennett & Hacker, 2007, p. 23) to the psychological domain. Thus, an account like S-representation is doomed to fail, the thought goes, because the claim that the brain (or any subpersonal aspect of a cognitive system) represents is nonsense.

Key to the S-representationalist response must be the idea that cognitive science is driving an empirically led change in our understanding of the concept of representation. No a priori constraints prohibit the application of representation to the subpersonal level, and by revealing commonalities between the personal

and subpersonal levels, cognitive science alters how we attribute representations. Even if the implicit rules for the use of representation may have once implied the necessity of agents, science has changed these rules because of what it has discovered about new domains. We suggest two ways of elaborating on this strategy.

First is the 'homuncular functionalist' approach (e.g., Dennett, 1975; Fodor, 1968; Lycan, 1991). Homuncular functionalism can be taken as one version of the 'technical' view of applying psychological predicates to the subpersonal level (following Figdor, 2018). Technical views state that scientific uses of psychological predicates at the subpersonal level are (typically) both literal *and* meaningful (contra Bennett & Hacker). However, psychological predicates do not refer to the same set of properties as the original domain—they pick out something different. Yet some significant similarity relation between the new and conventional referents is preserved. Homuncular functionalism postulates that cognitive capacities at the personal level can be explained by positing levels of sub-systems with capacities that are progressively less sophisticated than the levels above. This allows psychological terms to be attributed to the sub-personal level so long as they are increasingly 'less intelligent', and eventually bottom out in brute, non-psychological capacities. In effect, this means that no mereological fallacy is committed when attributing representation to the subpersonal level so long as such representations do not possess all the features of the capacity they explain, and are eventually decomposable into non-psychological parts. Put otherwise, any psychological or 'homuncular' implications at the subpersonal level are innocuous so long as the cognitive properties in question are progressively attenuated.

Homuncular functionalism (and other variants of the technical view) remain popular in the philosophy of mind and should be considered a viable strategy for combating a priori objections to the S-representation account. However, among its implications is that attributions of psychological predicates reflect a change in meaning. This is because psychological capacities ascribed to parts cannot be the same capacities as those at the personal level (remember, parts must be progressively stupider). A second way to elaborate the S-representationalist response that avoids this requires us to insist that representation is being used univocally at the personal and subpersonal level but that no fallacy is committed when we attribute the very same capacity of a whole to its parts (Figdor, 2018). This is because there is no a priori reason why the same properties cannot be found at multiple levels. Planets rotate and this is partly explained by the rotation of their atoms, with no change in the meaning of 'rotation'. The same can be said of psychological terms. Whilst a psychological predicate is so named because it originally described persons (as rotation originally described macroscopic objects), science shows how these terms capture patterns of behaviour in other domains (as it did for rotation at the molecular level). In other words, science applies the same psychological predicates to human agents and the subpersonal level because it has uncovered relevantly similar structures across the two domains. A version of this view is defended by Figdor

([2018](#)) who argues that formal models, in particular, provide grounds for attributing the capacities that we ascribe to human agents using psychological language to other domains, including the subpersonal level.[7]

There is no formal model for representation (though the structural correspondence condition for S-representation can be at least partially described using formalisms of isomorphism or homomorphism). However, the S-representation account does articulate a set of conditions that are intended to be realisable at the subpersonal level whilst capturing the functional character of a class of ordinary representation. This at least provides the basis for a qualitative analogy that resists the pull of the claim that the subpersonal level cannot instantiate the capacity to represent. This is the case even if such a capacity were quintessentially 'personal-level', because the S-representation account, in responding to scientific developments, shows how the functional profile of that capacity is recapitulated in the new domain. To borrow from Dennett ([2007](#)), in response to Bennett & Hacker; "It is an empirical fact, and a surprising one, that our brains—more particularly, *parts* of our brains—engage in processes that are *strikingly like* guessing, deciding, believing, jumping to conclusions, etc. And it is *enough* like these personal level behaviours to warrant stretching ordinary usage to cover it" ([2007, p. 86](#)).

A different flavour of a priori eliminativism is arguably found within strains of enactivism, insofar as they suggest the concept of cognitive representation misconstrues the fundamental ontological relationship between mind and world (e.g., Varela et al., 1991). Enactivist views on representation are often vague and heterogenous, and we cannot hope to dissect every variant of the enactive approach to representation here. Thus, we will settle for noting that at least some enactivist views make clear that they are concerned with the personal level relationship between an agent and the world (particularly perceptual objects), and not subpersonal mechanisms (e.g., see Noë, 2004; O'Regan, 2011). This paper is concerned with subpersonal mechanisms and so the consequences of the account for claims about the relationship between mind and world at the personal level remain, at worst, ambiguous. We also note that certain 'radical enactivist' approaches (Hutto & Myin, 2012) that do encompass the subpersonal level are principally concerned with the ability of representational theories to demonstrate how a mechanism may function in a way that makes content functionally relevant. We engage with this view in §3.3.

---

[7]For example, the Drift-Diffusion Model of two-choice decision-making (Ratcliff & McKoon, 2008) applies to the behaviour of both fruit flies and humans (Figdor, 2018). This model accounts for relatively automatic two-choice decisions in terms of accumulating information over time at a certain rate (the 'drift rate') until evidence for one option exceeds a prespecified threshold and a response to the decision-making task is given.

3.2   Function objections

The most straightforward way to undermine the S-representation account is to interrogate whether a mechanism meeting the functional conditions it establishes really qualifies as a representation (for discussion, see Facchin, 2021a; Gładziejewski & Miłkowski, 2017; Nirshberg & Shapiro, 2020). The clearest expression of this objection, to our knowledge, is offered by Facchin (2021a), who argues that S-representations cannot pass the job description challenge (JDC) because their functional profile is satisfied by so-called 'receptor representations' which paradigmatically fail the JDC.

The idea underlying receptor representations is that if an internal state of some system nomically covaries with some distal event, then that state represents the event. This has its origins in the notion of 'natural meaning' elucidated by Grice (1957), and paradigm cases such as 'smoke means fire' and 'spots on the face mean measles'. Applied to receptors we might intuit that a single neuron, for example, which fires in the presence of a stimulus represents that stimulus. Facchin joins others (e.g., Ramsey, 2007) in suggesting that receptors do not qualify as real representations. This is because, as basic 'causal mediators', their role is not sufficiently representation like, even if we complement the receptor notion with something like a Dretskean theory of content, which specifies that some receptors play an important 'indication' role for the systems in which they are embedded (Dretske, 1988). Moreover, many clear-cut cases of non-representation appear to meet the conditions of receptors, for example, states of a firing pin in a gun nomically covary with trigger position. Extending the label of 'representation' to encompass receptors would weaken the explanatory role of representation ascriptions, and invite a kind of 'pan-representationalism'. The S-representation account is, in part, motivated by the goal to provide representation ascriptions with a more explanatory significant role than the receptor notion allows.

The problem, Facchin claims, is that certain receptors meet the criteria for S-representation, as set out above. To be more precise, all receptors possess action-guiding structural correspondence with the target they are supposed to 'indicate' for a system, whilst some receptors can be made decouplable and sensitive to error. In brief, Facchin claims that receptors will always meet at least one kind of exploited structural correspondence given the dependence on temporal passage for indication; a system whose functioning depends on receptors to indicate an event is (at minimum) sensitive to the temporal relations holding among features of the receptor mechanism (e.g., the temporal relations holding among different lengths of a bimetallic strip within a thermostat), and the obtaining of a time-dependent structural similarity between the receptor and the target system (e.g., a set of environmental temperatures). For brevity, we cannot discuss Facchin's examples at length, so we will accept for the sake of discussion that these are bona fide counterexamples. But even granting these counterexamples, Facchin's argument can be responded to.

Several possible solutions stand out. The first solution notes that if the reasoning underpinning the function objection is sound, then the four conditions set out above cannot be sufficient to capture the functional profile of ordinary maps and models, and further work is required to identify the property or properties of S-representations that distinguish them from receptors. After all, the thought goes, if a type of representation is defined by a certain functional profile, then there must be *some* functional feature of maps and models that we've missed. Thus, the S-representationalist must identify a fifth condition that characterises such ordinary representations that can be mirrored in cognitive systems. Whilst we think this solution is plausible, without any current suggestions on the market, it puts the S-representationalist on the back foot.[8]

The second solution also acknowledges that a limited number of mechanisms we classify as receptors meet the conditions set out by the S-representation account but clarifies that the issue lies in the coarse-grain nature of the functional profile in most presentations. It is not *mere* structural correspondence or system-detectable error that is vital to the distinctive functional character of S-representation, but exactly how these broad-brush functional characteristics operate. Put differently, such descriptors are coarse-grained placeholders for a more detailed functional specification. Like the first solution, this forces the S-representationalist to reconsider how to formulate their account, but it has the benefit of a clearer path to achieving this. For instance, whilst Facchin (2021a) may be correct that even the most basic receptor instantiates a non-epiphenomenal structural correspondence with its target, this correspondence alone is clearly of a weak variety compared to the sort of correspondence that is supposed to be at work in, say, a cognitive map underlying counterfactual reasoning during prospective route planning. Of course, further work is required to define precisely what the difference-making correspondence is in the latter case, but the strategy is clear enough.

A variant of this second solution side steps any need to define different types of structural correspondence, or any other condition for S-representation, by instead claiming the difference between S-representations and (some) receptors is one of degree, not kind (cf. Nirshberg & Shapiro, 2020). What qualifies S-representations but not receptors for the job of representing is the richness or complexity of functional characteristics like structural correspondence. This may entail that representation is a graded notion with ambiguous cases, but many useful notions are graded, and the existence of borderline cases would not disqualify exemplars (Clark & Toribio, 1994). This response is developed by Rutar et al. (2022), in the context of the predictive processing framework, who argue that at least two conditions for S-representation are gradual features, by appealing to empirical evidence:

---

[8]Facchin discusses this solution but counters with two problems: (1) there is no obvious candidate for a fifth condition, and (2) the S-representation account, specifically, Gładziejewski's (2015) version, is already demanding. We agree the lack of candidates for a fifth condition is problematic but are less convinced by the second worry; the stringency of the S-representation account is a virtue, and besides, we cannot judge how much more demanding the S-representation account would become before any candidates have been identified.

structural similarity (in terms of the granularity of state space and the number of exploitable relations) and decouplability (in terms of how much neural structures depend on internal versus external stimulation, and the extent to which they are subject to precision weighting of prediction error). Certain developmental considerations support this response; if there is phylogenetic or ontogenetic continuity between some receptors and S-representations within cognitive systems, we might expect a great deal of overlap in their functional profile.

The third solution acknowledges that some receptors meet the conditions for S-representation but bites the bullet and accepts, despite any intuitive classification to the contrary, that we should treat these too as passing the JDC, and thus, as representations. Notice, however, that not *every* apparent receptor meets these conditions, e.g., single neurons firing in response to a stimulus. Indeed, Facchin's (2021a) examples of receptors which meet the functional profile for S-representation add several bells and whistles to standard examples. Even if we must widen the scope of JDC-passing mechanisms, we are far from trivialising representation. Put otherwise, if some receptors satisfy the same profile as S-representations *and* we allow those receptors to pass the JDC, then there is no fatal threat to the S-representation account. And indeed, some theorists have embraced the conclusion that some receptors pass the JDC (see Artiga, 2022; Morgan, 2014). In short, Facchin's argument only appears forceful if we refuse to grant some receptors representational status, which the S-representationalist is not obliged to do.

These promising solutions provide reasons to think the S-representation account has ways of escaping the function objection. Nonetheless, we believe this challenge should be welcomed to the extent that it forces proponents to consider the functional properties of S-representation more carefully. Even if opponents are satisfied that the S-representation account establishes a convincing functional profile for cognitive representation, they may still question whether it allows for a plausible story about semantic properties at the subpersonal level.

3.3   Content objections

Representations are *about* things—they possess 'content'. Representational content implies 'accuracy conditions' (encompassing truth, correctness, and other semantic measures of success). This means representations can get things 'right or wrong' about what they represent. In other words, representations are 'semantically evaluable'. Representational artefacts acquire their content, at least in part, via the norms surrounding their use by ordinary agents. For instance, a cartographic map represents a geographical region, at least in part, because an individual or community use the object to stand in for that geographical region. There are no agents to fix meaning in the case of subpersonal representation. This leads to a puzzle about how to make sense of the semantic properties of cognitive representations.

Following Lee (2018), there are two problems associated with content. The first is the 'hard problem of content' (HPC) (Hutto & Myin, 2012), which concerns why we should think of S-representations as possessing semantic properties in the first place; in other words, what justifies thinking S-representations are in the business of being *about* anything. The second is the 'content determination problem', which concerns the conditions for determining the content of a token representation; in other words, what makes a particular S-representation about x and not y?[9] We can ask analogous questions about familiar representations, like cartographic maps. On the one hand, we can ask what properties of maps allow for this type of artefact to function as content-bearing representations. On the other, we can ask what determines a particular map's content. Let's examine these two problems in turn.

Hutto & Myin (2012) introduce the HPC as a challenge for any naturalistic account of cognitive representation. The thought goes that for something to count as a genuine representation, we must account for its semantic properties in naturalistic terms. And yet, they conclude, there is no satisfactory account of content at the level of 'basic cognition'. Instead, content only appears with the emergence of intersubjective norms that provide standards for determining semantic properties (Hutto & Myin, 2017; cf. Zahnoun, 2021).

More recently, Segundo-Ortin & Hutto (2019) have challenged the S-representation account on similar grounds, suggesting proponents often presuppose but do not explain the origin of content (see also Hutto & Myin, 2017). They characterise the reasoning underpinning the S-representation account 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. (Segundo-Ortin & Hutto, 2019, p. 10)

However, they go on to conclude that whilst structural similarities might enable semantic evaluation, they are not themselves contentful:

> [I]t does not follow from the fact that we can make truth evaluable 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. (Segundo-Ortin & Hutto, 2019, p. 13)

---

[9] This partitioning into two sub-problems is also implicit in Segundo-Ortin & Hutto's (2019) critique of the S-representation account, where they suggest that concerns about content indeterminacy are a distraction from the more serious problem of how structural similarity can ground content in the first place.

In short, Segundo-Ortin & Hutto accept that structural similarities play a causal role in enabling successful cognition but insist this does not imply structural similarities are contentful.

Perhaps the most promising way to think about the role of content in the S-representation account is in terms of how it captures the distinctive relationship between a mechanism which meets the conditions for S-representation, a target item, and behavioural success. Segundo-Ortin & Hutto are correct that structural similarity is insufficient for content, but the S-representation account emphasises that mechanisms represent by virtue of the part they play in cognitive tasks. S-representations explain when the following conditions apply: (1) a system undertakes some task (e.g., navigating to a target location); (2) the outcome of that task depends on a mechanism with a component that structurally resembles task-relevant items (e.g., a cognitive map); (3) success depends on the degree of structural similarity between the mechanism component and those items (e.g., the topographical features of a rat's local environment). In other words, where an S-representation plays a causal role in realising a cognitive capacity, that capacity causally depends on the degree of structural correspondence between the mechanism and some target item(s). Given these features, the mechanism is described as functioning as a stand-in, surrogate or simulation of the target. Attributing accuracy conditions that are met if the required correspondence occurs thus captures a facet of the role of the mechanism (its representation-like function). Describing a cognitive map as accurate when it corresponds to the rat's environment (causing success) and inaccurate when it does not correspond to the rat's environment (causing failure) captures a distinctive and explanatorily relevant relationship between mechanism, target and behaviour.

The same story is true, we think, for the familiar representations which inspire the S-representation account; ascribing content to ordinary maps and models is explanatory because of the relationship between vehicle, target and behavioural success. Describing a cartographic map as accurate when it corresponds to the mountaineer's environment (causing success) and inaccurate when it does not correspond to the mountaineer's environment (causing failure) captures a crucial relationship between map, target and behaviour. If we are right, then content ascriptions are explanatory when behavioural success depends on an item playing a certain role, regardless of whether it is a conventional artefact or an evolved mechanism. The S-representationalist no more 'presupposes' content when explaining cognition by appealing to the exploitation of S-representations than we do when explaining the movement of tourists around London by appealing to the exploitation of an Underground map; in each case, accuracy conditions fall out of the fact that behavioural success depends on an item playing the role of a stand-in, surrogate or simulation for some target.

One might think the difference between the cognitive and ordinary map is that the consumer of the latter is a human agent and this is essential for bona fide content; therefore, so-called cognitive maps are not *real* representations. We

see no reason to assume agents are essential for content (see §3.1). However, even if content implied agency, we think cognitive science would be compelled to ascribe something content-like to mechanisms that met the conditions for S-representation. Let's call this 'schmontent'; when a mechanism meeting the criteria for S-representation sufficiently corresponds to an item such that it causes successful behaviour, all else being equal, we can call it 'schmaccurate', and 'schminaccurate' when it does not. If the anti-representationalist would be appeased by distinguishing content from schmontent, the S-representationalist should acquiesce, for the quarrel would transpire to be more-or-less terminological, and more-or-less everything that mattered about S-representations for scientific explanation would be preserved.

Segundo-Ortin & Hutto (2019) also raise valid concerns about the rush to draw an analogy between cognitive maps and ordinary maps based on a still somewhat incomplete understanding of the mechanisms involved (principally concerning whether 'forward sweeps' of activity in place cells implicated in anticipatory movement are used by the brain as surrogates for available routes). First, this objection amounts to advising caution about assigning representational function to cognitive maps whilst our mechanistic models remain impoverished. It does not demonstrate cognitive maps are *not* functioning as maps—only that we require more evidence. Second, cognitive maps are only one case of purported S-representation. Third, this objection illustrates the shift from more conceptual concerns about the possibility of mechanisms bearing subpersonal content, and towards productive concern for the relative empirical support for the existence and range of S-representations in the brain (see §4).[10]

Even if the hard problem is overcome, the S-representation account must still address the content determination problem. Broadly, the worry is that whilst it may be permissible to speak of S-representations bearing content, in the abstract, the S-representation account lacks a satisfactory story about how a token mechanism comes to represent the particular item it does. Moreover, many have worried the S-representation account implies a positive but implausible story about how content is fixed: mechanisms represent by virtue of bearing structural similarities to items in the world, the thought goes, so an S-representation must be about what it shares structure with. Unfortunately, structural correspondence is cheap. Whilst a hippocampal map might correspond to a rat's environment—which intuitively relates to its content—it also corresponds to very many other things in the universe too. If structural correspondence fixes content, then this leads to 'massive indeterminacy' (Sprevak, 2011, p. 671).

The content determination problem is best addressed by showing how the S-representation account is compatible with a story of content that does not lead to

---

[10]These objections may be mutually supporting. For instance, one's negative appraisal of the evidence for analogy between brain mechanisms and ordinary maps might support existing suspicions that nothing can naturalise content at the subpersonal level. Our adding precision to the debate by taxonomising objections does not imply the objections are not reciprocally reinforcing.

indeterminacy. There are broadly two ways of achieving this: 'hybrid' and 'task-oriented' approaches. Rather than siding with one or the other, we will provide an overview of both. This is especially appropriate given that the hybrid and task-oriented approaches may be integrated.

Many theories—from causal dependency theory to asymmetric dependency theory to teleosemantics—have been claimed to resolve the content determination problem for representationalism. Nevertheless, proponents of the S-representation account have sometimes formulated their theory, in part, as a response to the perceived inadequacy of these other theories to fully address the role and value of cognitive representation. For instance, Ramsey (2007, 2016) claims that theories like teleosemantics fail to identify why representations play a role in cognitive science in the first place; by themselves, traditional theories of content do not establish that anything functions in a recognisably representational manner. However, he also proposes that such theories do plausibly address the content determination problem. Hence, a promising strategy is to combine the S-representation account with a traditional theory of content to create a 'hybrid account'. Under this division of labour, the S-representation account specifies which cognitive mechanisms count as representations whilst, say, teleosemantics specifies the conditions that determine what a given representation is about.

A second approach to the content determination problem indicates that an account of content determination is already implicated by the causal role of S-representations in guiding the actions of a system (for related discussion, see Lee, 2021; Piccinini, 2022). Whilst the minutiae vary between theorists, proponents of such task-oriented approaches are often careful to distinguish between two kinds of representational objects, already implicit in much of our discussion so far, which we will refer to here as the 'target' and 'content' of the representation. The former are, roughly, those conditions the representation is 'applied to' (Gładziejewski, 2015, p. 80) or 'used to deal with' (Godfrey-Smith, 2006, p. 58). The latter are, roughly, those conditions that must occur for the capacity that depends on the mechanism to succeed. This can be thought of as the difference between what a representation must refer to for behavioural success to occur given the actual conditions of the task versus what a representation does refer to given how its structure guides the system. For example, the target of a cognitive map might be the structure of a novel maze a rat is currently navigating whilst its content refers to the structure of the maze it was previously trained on and which the map developed in response to. The accuracy of a given S-representation is a function of the overlap between target and content.

We believe much of the task-oriented story aligns with the response given to the HPC above. However, we also note that the task-oriented and hybrid approaches are not necessarily exclusive. For instance, one may hold that to properly understand what a representation is being applied to—its target—one requires a naturalistic account of what task the system is performing, or what the mechanisms 'proper function' is, and this is something only an account such as teleose-

mantics is equipped to address (for a thorough account which combines elements of both approaches, see Shea, 2018). In short, whether adopting the hybrid or task-oriented approach, or some combination, the S-representationalist has the resources to construct an account of content determination that does not lead to massive indeterminacy.

In closing our discussion of the first three ('conceptual') objections, we should note an emerging theme, namely, that allowing for the possibility of S-representation at the subpersonal level invites the discovery of interesting similarities between aspects of familiar problem-solving and subpersonal cognition. We do not claim the S-representation account wholly reflects the folk usage of 'representation' (as if it was well-defined, to begin with). Indeed, following a broadly Quinean tradition, we take it that much philosophy is not in the business of uncovering the true meaning of clear-cut terms but rather refining how we can or should use a term for an area of discourse that requires high levels of precision. The point of the S-representation account, in our eyes, is not that we are compelled to use representational language *de rigeur* but that when the four conditions for S-representation (structural correspondence, action-guidance, decouplability, system detectable error) are met by a cognitive mechanism, the representational label effectively captures its functional role, much like labelling the heart as a pump is effective because of the strong resemblance between its activity and those of ordinary pumps. In this way, conceptual precision is aided through the redeployment of folk terms, in turn, refining what those terms mean within a particular (namely, scientific) context.

4   Empirical objections

Suppose we accept both that subpersonal representation is possible, and that S-representations, if they exist, function as representations with sufficiently determinate content. The anti-representationalist may nevertheless question whether S-representations belong in our best scientific theories. If S-representations do not feature in our best scientific theories, then the mere philosophical coherence of the account is not relevant to explaining cognition.

There are two subtly distinct ways in which critics can argue the S-representation account falters in this regard. First, one can argue that a non-S-representational theory of cognition is better than an S-representational theory of cognition, where everyone more-or-less agrees on which theories are representational and which are not. Here, the disagreement concerns which theory is best and not which theories posit S-representations. Second, one can argue that our best theory of cognition, thought by opponents to involve S-representations, features nothing of the sort. Here, the disagreement concerns how to interpret the posits or constructs of our best theory and not which theory is best. There is an overlap between these objections, as we will see, but the delineation adds precision. The distinction also reflects contours in the literature,

where independent arguments have been offered for the superiority of a theory that is widely recognised as non-representational or for a non-representational interpretation of a more ambiguous theory. The first line of thought will be explored next as 'the best theory objection', and the second in the following section as 'the interpretation objection'.

4.1   The best theory objection

Some anti-representationalists are less concerned with the conceptual lucidity of cognitive representation and more with whether representation features in our best scientific theories. Here we consider this line of argument, concentrating on one especially prominent way in which the discussion has developed. If a non-representational theory offers a superior explanation of cognition, then via inference to the best explanation, an opponent of S-representation could conclude that even if the existence of S-representations is plausible, they are not in fact instantiated by cognitive systems. (A softer approach might conclude only that they are not useful for understanding real cognitive systems). To illustrate this line of argument, we will introduce a version of 'radical embodied cognition theory' as a case study, which draws on Dynamical Systems Theory (DST) (broadly based on the account given by Chemero, 2009, sec. 3). We will first present the reasons given for supposing that this anti-representational framework is a genuine alternative to representational theories. One popular reason for dismissing this version of embodied cognition as a genuinely explanatory theory—the question of 'representation-hunger'—will be considered but found not to be decisive. In response, we will argue that the radically embodied view, which uses DST as a holistic paradigm for explaining cognition, is unpersuasive as an alternative to representationalism, and that the direction of travel in the scientific and philosophical literature shows that DST is better understood as a useful tool for modelling neural dynamics within a paradigm that is amenable towards S-representations.

Some anti-representationalists, such as Chemero (2009), have argued that DST provides a framework for understanding cognition that treats the brain, body, and environment as a complete system whose behaviour can be modelled mathematically and without appeal to representations. DST predicts and explains how the states of a cognitive system will evolve through time. Notably, the complex reciprocal relationships between neural states and states of the environment are not characterised by Chemero (2009, sec. 2) in terms of representation, but as captured by a set of mathematical functions that govern the dynamics of the system.

Parallels are drawn between cognitive systems and the Watt governor (Gelder, 1995). The Watt governor is a control system that acts to regulate the speed of an engine by limiting the amount of fuel it receives; as the speed of the engine increases, the centrifugal force it generates is harnessed by a pair of weighted arms to gradually close a valve to slow the consumption of fuel. Thus, the governor maintains a constant engine speed as the entire system is engineered to remain

in a dynamic equilibrium. It is possible to provide a representational description of the governor's operation by supposing that the angle of the arms represents the speed of the engine for the system, and that representation is 'used' by the system to control the aperture of the fuel valve.[11] However, the thought goes, the dynamical (non-representational) explanation is more elegant and predictive. For any conditions, the dynamical equations allow us to calculate precisely how the system will evolve.

Despite sceptical claims that cognition is not the kind of phenomenon that can be explained using a set of dynamical equations, there are good reasons to suppose that this version of radical embodied cognition forms the basis for a progressive research program, in Lakatos' sense (Chemero, 2009, p. 207). The Haken-Kelso-Bunz (HKB) model, for example, has provided a fruitful experimental paradigm. HKB uses the notion of an oscillator—a system that is stable but moves repeatedly between two or more states in a regular pattern—to characterise aspects of motor control such as walking. HKB can also be extended to be applied to coordination problems (e.g., limb movements, Kugler et al., 1980).

A common challenge raised against DST, however, is that despite providing an interesting angle for understanding automatic motor control, it cannot explain so-called 'representation-hungry' tasks, such as abstract problem-solving, requiring detailed visualisation and subsequent mental manipulation in the imagination (e.g., calculating a line of chess moves, solving a Rubik's cube, or considering which piece of art will best complement your interior design palette). Nevertheless, DST has made promising developments in answering this challenge. Stephen et al. (2009), for instance, asked experimental subjects to solve a gear-system problem in which an arrangement of gears is displayed, and the subjects are asked to infer the direction of rotation of a particular gear based on information about the rotation of one other gear in the arrangement. Novices in the task use an inefficient and cognitively demanding strategy. However, following some experience, subjects develop a more efficient strategy based on an abstraction (counting the gears). The authors found that they could accurately predict this change in cognitive strategy using a dynamical analysis of action during problem-solving. By treating the subject as instantiating a Lorenz attractor—a form of unstable attractor that characterises a system which transitions between several well-defined states—they could map the overall entropy of the subjects' actions to this well-understood trajectory through the state-space of the overarching system. In this way, dynamical systems theorists use behavioural analyses to discover underlying patterns that can be captured by relatively simple mathematical models of our cognitive processes. Similar studies use mouse tracking to collect data on hand movements during decision-making to understand subconscious biases (McKinstry et al., 2008) and the structure of language (Spivey et al., 2005).

---

[11]The representational account of the Watt governor is usually expressed in terms of a covariance relation and does not meet the criteria for S-representation (cf. Bechtel, 1998).

Citing examples such as these, some advocates of radical embodied cognition argue that DST offers at least some of the tools needed to form a credible alternative to theories of subpersonal cognition that appeal to representational mechanisms. In addition to the success of empirical work and the progressive nature of this research, there appear to be broader theoretical advantages. Whilst classical computational theories focus on intracranial processes, DST appeals to facts about the entire brain-body-environment system to provide accurate and robust models. In doing so, it avoids the (apparent) decompositional strategy of some mechanistic theorists, treating cognition as a phenomenon that cannot be decomposed into discrete subsystems. Dynamics are often chaotic, with small changes in the structure of a system or its initial conditions leading to consequences that are strictly unpredictable. Thus, theoretical commitments may lead us to conclude that representational mechanisms are not part of our best explanations.

As radical embodied cognition illustrates, the best theory objection focuses attention on the relative virtues of competing explanatory strategies (pluralism notwithstanding). The objection, taken alone, allows that S-representation is a coherent and potentially explanatorily useful concept, objecting only that a non-representational theory (here, using tools of DST) is preferable. In §2 we clarified that the concept of S-representation is not a theory of cognition, but refers to a type of posit or construct that may be invoked by different theories. The acceptance of S-representation as a coherent and potentially-useful concept would be a significant concession to its defenders. In turn, those defending the empirical utility of S-representations may feel unthreatened by 'best theory' objections, believing that explanatory strategies involving representation are superior to those that do not, hence their hard work to elucidate the conceptual details of the posit or construct. Their confidence may be justified given the continued dominance of model-based theories in cognitive science.

The success or failure of the best theory objection will ultimately depend on future scientific developments. In this sense, those involved in the debate can put aside their differences while promoting the empirical work they believe to be most promising. It is worth noting at this point, however, that a significant volume of work on DST embraces the notion of representation and seeks to use system dynamics to unify computational explanations with underlying neural dynamics (e.g., Cisek, 2007; Clark, 2008; Eliasmith & Anderson, 2003; Piccinini, 2022). In contrast to a more rigidly demarcated pluralism, which denies representation a role within DST but allows for the coexistence of different frameworks, this more 'inter-theoretic' approach suggests that DST and S-representation can operate as part of the same explanatory package.

One way this might work, in brief, is that representational models make sense of the functional mechanisms within the brain, with DST elucidating how S-representations are instantiated by neural dynamics and, at a greater spatio-temporal scale, the interaction between whole brain-body-environment systems. At the neural level, dynamics may enable us to understand how the behaviour of

a single cell or set of cells, whose states may vary in many different ways, can be constrained by their organisation to encode discrete computational vehicles. In this way, these analogue systems might even be taken to instantiate classical computational structures (Eliasmith & Anderson, 2003, p. 55). An example of this is the oculomotor integrator, as analysed by Shagrir (2012). The speed and direction of saccades are transformed by this small set of neurons into signals used by the oculomotor system to remember the new position of the eye. Thus, the integrator circuit enables certain inputs to represent these properties of the eye, and the output represents the new position and is remembered for later use. Plausibly, the entire subsystem encodes an S-representation of possible changes in eye position which is realised by the particular dynamics of the neural activity.

Higher-level analysis of brain dynamics has primarily been used to motivate a more 'action-oriented' understanding of cognitive representation (Cisek, 2007; Remington et al., 2018). This approach rejects the notion that representations instantiate an agent-neutral model of the world, but rather represent 'affordances'. For example, rather than representing a water bottle as 'being 2 feet away', a water bottle may instead be represented as being 'reachable'. Affordances such as reachability are dependent on the embodiment of the agent and its capacities. This approach to content is not incompatible with the S-representation account, which stresses the importance of action guidance. Schöner (2019), for instance, offers a theory-neutral account of how dynamical and representational explanations can be effectively bridged that emphasises features such as decoupleability, stable structural relationships, and system-detectable error. Though a complete interpretation of Schöner's approach to representation through an S-representational lens must wait, we can begin to see how such features are suggestive of action-guiding map- or model-like structures i.e., candidate S-representations.

In brief, Schöner (2019) illustrates how the dynamics of excitation and inhibition of entire neural populations can realise intentional mental states, together with their conditions of satisfaction (CoS). Similar to the oculomotor integrator, the dynamical relationships between these populations instantiate a structure that facilitates the use of explanations involving S-representations. In Schöner's example, one neural population (A) is taken to instantiate a particular intentional state. This may play any number of roles within the neural system, being activated by top-down, horizontal, or bottom-up connections originating from other neural populations. Another neural population (B) functions to inhibit the activation of A. B's inputs are taken to covary with the CoS for the intentional state instantiated by A. Thus, when A and B are active, the system is actively representing a particular state of affairs, and that state of affairs obtains. Any errors represented by A will be detected and processed as residual activation flows as a result of imperfect inhibition by B.

Thus, the possibility of mapping the activation of a neural population encoding one intentional state to another which correlates with the CoS for that state enables a basic S-representational interpretation. This is strengthened by further structural

relationships, however, such as the inhibitory coupling between the two; once the prediction of success encoded by the intention is fulfilled by the activation of the CoS population (B), the intentional population (A) is inhibited, and consequently, there is no error to be detected. In this way, inhibitory dynamics can be interpreted as instantiating system-detectable error, coherent with an S-representational understanding of the system.

The upshot of such bridging analyses is that we need not adopt a binary approach when considering the explanatory success of S-representational theories and radical embodiment based on DST. There are good reasons to suppose that both strategies can and should be taken together to bring about a full understanding of cognition. Rather than treating DST as forming the basis for a complete explanatory paradigm, it should be (and has been) adopted as a useful tool for modelling the dynamics of neural populations in a way that provides further support to representational explanations of cognition. The broader lesson for thinking about the best theory objection is that we must carefully assess whether the supposed rival to a representational theory really does embody a competitor.

Even if the pro- and anti-representationalist can agree over which theory in cognitive science is best, the latter may still insist that S-representations are empirically inert because we are mistaken in supposing that our best theory features S-representations to begin with. This brings us to the interpretation objection.

4.2   The interpretation objection

If the S-representation account provides a convincing conceptual foundation for subpersonal representation, and all parties agree on our best theory of cognition, disagreement may still be found over whether this theory posits S-representations. Whilst the best theory objection is chiefly concerned with which theory of cognition is best, the interpretation objection is concerned with how to construe our best theory's constituents.

At this stage, more overlaps in our taxonomy become evident; if one is convinced by a priori objections to subpersonal representation (see §3.2), then one will be driven to a non-representational interpretation of a theory. However, 'a priori' and 'interpretation' objections still come apart because one can contest the representational credentials of a theory without thinking subpersonal representation is a category mistake—as evidenced, for instance, by Ramsey's (2007) analysis of representation in different theories of cognition. We should also acknowledge potential ambiguity over whether a disagreement concerns competition between what our best theory is and what our best theory entails. Two interpretations of predictive processing (PP), for instance, might diverge so much that we no longer count them as the same theory. Regardless, we separate the best theory and interpretation objections for added precision, noting the representationalist and anti-representationalist may agree on a great many things concerning the best meth-

ods, models, evidence and descriptive tools for understanding cognition, but still disagree over whether these imply S-representation.

The interpretation objection can be illustrated with reference to predictive processing. Very briefly, PP offers a theory of cognition in which the brain is organised to minimise error in its own internally generated (top-down) predictions of the incoming (bottom-up) sensory input (Clark, 2016; for an introduction, see Friston, 2009; Hohwy, 2013). In orthodox formulations, predictions are measured against sensory input, producing prediction errors that are used to update a multi-level 'generative model' which determines future predictions (e.g., Clark, 2016). This is achieved by encoding prior expectations ('priors') about sensory input, pitched at multiple spatial and temporal scales spread across a processing hierarchy. At higher levels in the generative model, priors consist of a set of 'hypotheses' that reflect expectations about the hidden (worldly) causes of stimuli, and an assignment of probability to those hypotheses. According to some versions, hypotheses are selected (assigned a 'posterior value') as a function of their prior probability plus their 'likelihood'—the probability that the state of affairs captured in the hypothesis would cause the received sensory input, were it true—in approximate accordance with Bayes' theorem.

The S-representation account has been used to identify the representational posits or constructs of PP (Gładziejewski, 2016; Kiefer & Hohwy, 2018; Wiese, 2017). One noteworthy claim found in representational treatments of PP is that the generative model functions as a kind of S-representation (or set of nested S-representations). The general idea is that for the brain to endogenously generate a prediction of the sensory signal it must embody a causal-probabilistic structure that maps onto the hidden worldly causes of stimuli, accomplished through encoding a multi-level network of updatable priors. Gładziejewski (2016) offers such a view when he writes,

> [C]ognitive systems navigate their actions through the use of a sort of causal–probabilistic "maps" of the world. These maps play the role of representations within the theory. Specifically, this map-like role is played by the generative models. It is generative models that, similarly to maps, constitute action-guiding, detachable, structural representations that afford representational error detection. (2016, p. 569)

In response, opponents have sought to undermine PP's representational credentials. There are different ways to attempt this Kirchhoff & Robertson (2018). However, one common strategy for combating representational interpretations of PP is to downplay the necessity of understanding notions like priors and generative models as representational, and instead recommend we understand such posits or constructs in terms of hardwired constraints, biases, sensory attunement and other

'ecological' notions.[12] For instance, writing about related Bayesian approaches to visual perception from an ecological point of view, Orlandi writes,

> [P]riors do not look like contentful states. They do not need to have accuracy conditions to perform their function. They rather look like built-in or evolved functional features of perception that skew visual networks toward certain configurations. (2014, p. 82)

Engaging with Orlandi's thesis, Downey claims,

> The concepts "prior" and "likelihood" are better understood as referring to mechanisms which pre-dispose brains to configure themselves into specific organisational patterns in response to environmental stimulation [...] Biases do not pass the job description challenge, and so we have no reason to treat them in terms of representation. (2018, p. 5121)

Hutto offers a related interpretation of PP, from a 'radical enactivist' approach (reworking a passage from Clark, 2016, p. 27):

> Nothing here requires [the brain] to engage in processes of ... [contentful] prediction or expectation. All that matters is that ... [its] systems be able to [anticipate and be adjusted by sensory perturbations] in ways that make the most of whatever regularities ... [to which it is attuned, because such attunement has] ... proven useful ... [in response to such regularities in the past]. (Hutto, 2018, p. 2456. Original parentheses.)

Recently Facchin (2021b) has suggested that generative models in PP systems do not satisfy the conditions for S-representations, as they either do not satisfy the condition of distality[13], or they do not satisfy the condition of exploitable structural similarity. According to Facchin, PP systems qua computational networks can feasibly represent only the proximal patterns of input activation, as it is only to this proximal activation that the generative model bears an exploitable structural correspondence. Thus, PP systems cannot be interpreted as using S-representations. The general lesson of these perspectives is the possibility of an interpretation of PP that replaces talk of internal models bearing representational content with ecological concepts.[14]

---

[12]Though these arguments do not always single out the S-representation account, they are offered in the context of a literature which does interpret the posits or constructs of PP as S-representations (e.g., Gładziejewski, 2016).

[13]Note that this is not a condition included in our definition of S-representations.

[14]A somewhat different take is offered by Kirchhoff & Robertson (2018) who target a particular representationalist interpretation of PP that centres Kullback-Leibler divergence between posterior generative and approximate recognition densities which the authors think relies on a relation of covariance, and covariance does not constitute content (Hutto & Myin, 2012). We think the authors err in emphasising the role of generalised synchrony (mutual information) in the wider mechanism (meeting the criteria for S-representation) responsible for behavioural success (see

A satisfactory response to any 'interpretation objection' requires closely engaging with the particular details of the scientific theory in question—such as the constructs of PP—and how best to interpret them (and here we are not invested in defending a representational interpretation of PP, in particular). There are, however, some general considerations to bear in mind when developing a defence of S-representation. One important consideration is that the interpretation objection, like the best theory objection, is not itself an objection to the S-representation account *qua* an account of how cognitive systems might instantiate representing mechanisms, in principle. As noted above, it is possible to maintain the coherency of the S-representation account but object to its theoretical applicability. Moreover, the inability of PP (or any other theory) to provide a convincing example of S-representation only threatens the applicability of the account if (a) PP is our best theory, and (b) it is exhaustive i.e., there is no room for theories which do provide effective examples of S-representation alongside PP (this is most obviously possible if PP explains only a subset of cognitive phenomena). Like the best theory objection, proponents should welcome any transition from assaults on the very possibility of S-representation to a discussion of how widespread S-representations are in reality.

A second consideration is that whilst the interpretation objection relies on evaluating the posits or constructs of a given theory—and in this sense is an 'empirical' concern—the quotations above demonstrate that theorists can and do disagree over how best to characterise fundamental theoretical contents; they disagree over whether a posit or construct—such as a system of priors—should be interpreted as a representation (or even as existing at all, following Facchin, 2021a!). For such cases, we identify three options for those defending the empirical relevance of the S-representation account.

First, one can argue that disagreement is purely terminological. For instance, proponents can argue that a mechanism which predisposes a cognitive system to configure itself into specific organisational patterns in response to stimulation—once the details are hashed out—is just an S-representation without the label (for general consideration of the representation debate as purely terminological, see Haselager et al., 2003). So long as the four functional criteria might reasonably be attributed to a mechanism, that mechanism is an S-representation, regardless of how else it is described. However, the extent to which this consideration provides a plausible rebuttal to anti-representationalists is questionable. PP has spawned a quickly evolving debate over what exactly the theory is committed to, and diverging interpretations plausibly reflect different (and mutually exclusive) outcomes.

Second, one can acknowledge that the sorts of interpretations offered in the quotations above do reflect genuine alternatives to a representational interpreta-

---

§3.2). Regardless, it's worth noting the authors do not question the conceptual foundations of S-representation, only the capacity of PP to provide a convincing example. Moreover, they recognise their objection only targets a version of a representationalist interpretation of PP and is not a generalised argument against representationalism.

tion but maintain that the representationalist interpretation of PP is superior. We will not settle the debate here but note that (a) within the philosophy of cognitive science literature, at the very least, the jury is out, with some recent literature defending a representationalist interpretation, and (b) within practising cognitive science, talk of model-like mechanisms remains ubiquitous. Therefore, the proponent of the relevance of the S-representation account for cognitive science is, in our estimation, on a firm footing.

Third, one can argue that the sorts of interpretations offered in the quotations above do reflect genuine alternatives to a representational interpretation, however, such interpretations operate at a different (non-mutually exclusive) level of description. The representationalist does not claim representational descriptions are necessary; one can always explain a cognitive phenomenon in more brute 'ecological' terms. The claim is that representational language delivers some explanatory grip, from a certain level of description. For instance, we might describe a mechanism either as a system that anticipates and adjusts itself by sensory perturbations in ways that exploit regularities in its history *and* as an internal representation. As a general rule, the translatability of representational descriptions into non-representational descriptions does not imply the former are false. In the case of an emerging theory like PP, we may need to wait and see the extent to which progress depends upon understanding systems of updatable priors as action-guiding, decouplable, and error-sensitive models (for discussion on the early developmental stage of PP and the caution this engenders, see Dolega, 2017).

Finally, it is worth mentioning the possibility that PP provides less of a complete theory and more of a mechanistic schema (Machamer et al., 2000), in other words, it supplies abstract specification of a mechanism (or set of mechanisms) which must be 'filled in' by a more complete mechanistic account (for related discussion see Gładziejewski, 2019; Miłkowski & Litwin, 2022). If correct, then this may help diagnose the ambiguity of what, exactly, PP is committed to: in its most generic sense, PP does not provide enough mechanistic details to determine whether mechanisms meeting something like the S-representation profile are required. Regardless, we bet that the persistence of model talk within PP (and across cognitive science) indicates the continued relevance of S-representation for the foreseeable future.

5   Conclusion

There is growing consensus that the S-representation account offers the most relevant and promising account of representation for contemporary cognitive science. However, there are several types of objections, and these ought to be distinguished. Whilst this paper focused on reviewing objections (and their responses) to the S-representation account, the novel taxonomy of objections we offered will prove useful when assessing criticisms of representationalism more generally. Locating an anti-representationalist argument in relation to the categories we set out of-

fers non-partisan benefits. For the representationalist and anti-representationalist alike, it encourages a refined understanding of logically independent arguments. In turn, it helps identify the relationship between different criticisms, for example, diagnosing shared argumentative strategies among different expressions of the 'a priori' objection or distinguishing between purely conceptual and strictly empirical concerns. The protracted representation wars have often suffered a lack of precision which has led to confusion and misunderstanding; representationalism and its repudiation have meant many different things with varying implications for cognitive science. Improved clarity should be welcomed by all participants.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology, 14*(1), 43–64. https://doi.org/10.1080/09515080120033571

Artiga, M. (2022). Strong liberal representationalism. *Phenomenology and the Cognitive Sciences, 21*(3), 645–667. https://doi.org/10.1007/s11097-020-09720-z

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609. https://doi.org/10.1017/S0140525X99002149

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22*(3), 295–317. https://doi.org/10.1207/s15516709cog2203_2

Bennett, M., & Hacker, P. (2007). Selections from philosophical foundations of neuroscience. In M. Bennett, D. Dennett, P. Hacker, & J. Searle (Eds.), *Neuroscience and philosophy: Brain, mind and language* (pp. 3–48). Columbia University Press.

Boone, W., & Piccinini, G. (2015). The cognitive neuroscience revolution. *Synthese, 193*(5), 1509–1534. https://doi.org/10.1007/s11229-015-0783-4

Brigandt, I. (2013). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 477–492. https://doi.org/10.1016/j.shpsc.2013.06.002

Chemero, A. (2009). *Radical embodied cognitive science.* MIT Press. https://doi.org/10.7551/mitpress/8367.001.0001

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 362*(1485), 1585–1599. https://doi.org/10.1098/rstb.2007.2054

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* MIT Press. https://doi.org/10.7551/mitpress/1552.001.0001

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* Oxford University Press.

Clark, A. (2015a). Predicting peace: The end of the representation wars—A reply to Michael Madary. In T. Metzinger & J. M. Windt (Eds.), *Open MIND: 7(R).* MIND Group. https://doi.org/10.15502/9783958570979

Clark, A. (2015b). Radical predictive processing. *The Southern Journal of Philosophy, 53*, 3–27. https://doi.org/10.1111/sjp.12120

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind.* Oxford University Press.

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese, 101*(3), 401–431. https://doi.org/10.1007/BF01063896

Cummins, R. (1996). *Representations, targets, and attitudes.* MIT Press. https://doi.org/10.7551/mitpress/5887.001.0001

Dale, R., Dietrich, E., & Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive Science, 33*(5), 739–742. https://doi.org/10.1111/j.1551-6709.2009.01042.x

Dennett, D. (1975). Why the law of effect will not go away. *Journal for the Theory of Social Behaviour, 5*(2), 169–188. https://doi.org/10.1111/j.1468-5914.1975.tb00350.x

Dennett, D. (2007). Philosophy as naive anthropology: Commment on bennett and hacker. In M. Bennett, D. Dennett, P. Hacker, & J. Searle (Eds.), *Neuroscience and philosophy: Brain, mind and language* (pp. 73–94). Columbia University Press.

Dolega, K. (2017). Moderate predictive processing. In T. K. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing.* MIND Group. https://doi.org/10.15502/9783958573116

Downey, A. (2018). Predictive processing and the representation wars: A victory for the eliminativist (via fictionalism). *Synthese, 195*(12), 5115–5139. https://doi.org/10.1007/s11229-017-1442-8

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

Duijn, M. van, Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior*, *14*(2), 157–170. https://doi.org/10.1177/105971230601400207

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT Press.

Facchin, M. (2021a). Predictive processing and anti-representationalism. *Synthese*, *199*(3), 11609–11642. https://doi.org/10.1007/s11229-021-03304-3

Facchin, M. (2021b). Structural representations do not meet the job description challenge. *Synthese*, *199*, 5479–5508. https://doi.org/10.1007/s11229-021-03032-8

Figdor, C. (2018). *Pieces of mind: The proper domain of psychological predicates*. Oxford University Press. https://doi.org/10.1093/oso/9780198809524.001.0001

Fodor, J. (1968). *Psychological explanation*. MIT Press.

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Gelder, T. van. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, *92*(7), 345–381. https://doi.org/10.2307/2941061

Gładziejewski, P. (2015). Explaining cognitive phenomena with internal representations: A mechanistic perspective. *Studies in Logic, Grammar and Rhetoric*, *40*(53), 63–90. https://doi.org/10.1515/slgr-2015-0004

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582. https://doi.org/10.1007/s11229-015-0762-9

Gładziejewski, P. (2019). Mechanistic unity of the predictive mind. *Theory & Psychology*, *29*(5), 657–675. https://doi.org/10.1177/0959354319866258

Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: Causally relevant and different from detectors. *Biology & Philosophy*, *32*(3), 337–355. https://doi.org/10.1007/s10539-017-9562-6

Godfrey-Smith, P. (2006). Mental representation, naturalism, and teleosemantics. In D. Papineau & G. Macdonald (Eds.), *Teleosemantics* (pp. 42–68). Oxford University Press.

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*(3), 377–388. https://doi.org/10.2307/2182440

Haselager, P., Groot, A., & Rappard, H. (2003). Representationalism vs. Anti-representationalism: A debate for the sake of appearance. *Philosophical Psychology*, *16*(1), 5–24. https://doi.org/10.1080/0951508032000067761

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Hutto, D. (2013). Exorcising action oriented representations: Ridding cognitive science of its Nazgûl. *Adaptive Behavior*, *21*(3), 142–150. https://doi.org/10.1177/1059712313482684

Hutto, D. (2018). Getting into predictive processing's great guessing game: Bootstrap heaven or hell? *Synthese*, *195*, 2445–2458. https://doi.org/10.1007/s11229-017-1385-0

Hutto, D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press. https://doi.org/10.7551/mitpress/9780262018548.001.0001

Hutto, D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. MIT Press. https://doi.org/10.7551/mitpress/9780262036115.001.0001

Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, *195*(6), 2387–2415. https://doi.org/10.1007/s11229-017-1435-7

Kirchhoff, M. D., & Robertson, I. (2018). Enactivism and predictive processing: A non-representational view. *Philosophical Explorations*, *21*(2), 264–281. https://doi.org/10.1080/13869795.2018.1477983

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in Motor Behavior* (pp. 3–47). North Holland. https://doi.org/10.1016/S0166-4115(08)61936-6

Lee, J. (2018). Structural representation and the two problems of content. *Mind & Language*, *34*(5), 606–626. https://doi.org/10.1111/mila.12224

Lee, J. (2021). Rise of the swamp creatures: Reflections on a mechanistic approach to content. *Philosophical Psychology*, *34*(6), 805–828,. https://doi.org/10.1080/09515089.2021.1918658

Lycan, W. (1991). Homuncular functionalism meet PDP. In W. Ramsey, S. Stitch, & D. Rumelhart (Eds.), *Philosophy and connectionist theory* (pp. 259–286). Lawrence Erlbaum.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, *19*(1), 22–24. https://doi.org/10.1111/j.1467-9280.2008.02041.x

Miłkowski, M., & Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, *200*(6), 462. https://doi.org/10.1007/s11229-022-03891-9

Morgan, A. (2014). Representations gone mental. *Synthese*, *191*(2), 213–244. https://doi.org/10.1007/s11229-013-0328-7

Nirshberg, G., & Shapiro, L. (2020). Structural and indicator representations: A difference in degree, not in kind. *Synthese*, *198*, 7647–7664. https://doi.org/10.1007/s11229-020-02537-y

Noë, A. (2004). Action in perception. In *MIT Press*. 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* (pp. 1–20). Elsevier. https://doi.org/10.1016/B978-008044394-2/50004-X

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, *51*(1), 78–109. https://doi.org/10.1016/0014-4886(76)90055-8

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. https://doi.org/10.1016/0006-8993(71)90358-1

O'Regan, J. K. (2011). *Why red doesn't sound like a bell: Understanding the feel of consciousness.* Oxford University Press.

Orlandi, N. (2014). *The innocent eye: Why vision is not a cognitive process.* Oxford University Press.

Piccinini, G. (2020). *Neurocognitive mechanisms: Explaining biological cognition.* Oxford University Press. https://doi.org/10.1093/oso/9780198866282.001.0001

Piccinini, G. (2022). Situated neural representations: Solving the problems of content. *Frontiers in Neurorobotics*, *16*, 846979. https://doi.org/10.3389/fnbot.2022.846979

Ramsey, W. M. (2007). *Representation reconsidered.* Cambridge University Press. https://doi.org/10.1017/CBO9780511597954

Ramsey, W. M. (2016). Untangling two questions about mental representation. *New Ideas in Psychology*, *40*(A), 3–12. https://doi.org/10.1016/j.newideapsych.2015.01.004

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Remington, E. D., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics. *Neuron*, *98*(5), 1005–1019. https://doi.org/10.1016/j.neuron.2018.05.020

Rutar, D., Wiese, W., & Kwisthout, J. (2022). From representations in predictive processing to degrees of representational features. *Minds and Machines*, *32*, 461–484. https://doi.org/10.1007/s11023-022-09599-6

Ryle, G. (1949). *The concept of mind.* Hutchinson.

Schöner, G. (2019). The dynamics of neural populations capture the laws of the mind. *Topics in Cognitive Science*, *12*, 1257–1271. https://doi.org/10.1111/tops.12453

Segundo-Ortin, M., & Hutto, D. D. (2019). Similarity-based cognition: Radical enactivism meets cognitive neuroscience. *Synthese, 1*, *98*, 5–23. https://doi.org/10.1007/s11229-019-02505-1

Shagrir, O. (2012). Structural representations and the brain. *The British Journal for the Philosophy of Science*, *63*(3), 519–545. https://doi.org/10.1093/bjps/axr038

Shea, N. (2018). *Representation in cognitive science.* Oxford University Press. https://doi.org/10.1093/oso/9780198812883.001.0001

Sims, M., & Pezzulo, G. (2021). Modelling ourselves: What the free energy principle reveals about our implicit notions of representation. *Synthese*, *199*(3), 7801–7833. https://doi.org/10.1007/s11229-021-03140-5

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Sprevak, M. (2011). Representation reconsidered by William M. Ramsey. *British Journal for the Philosophy of Science*, *62*(3), 669–675. https://doi.org/10.1093/bjps/axr022

Steiner, P. (2014). Enacting anti-representationalism. The scope and the limits of enactive critiques of representationalism. *Avant: Trends in Interdisciplinary Studies*, *5*(2), 43–86. https://doi.org/10.26913/50202014.0109.0003

Stephen, D., Dixon, J., & Isenhower, R. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1811–1832. https://doi.org/10.1037/a0014510

Thomson, E., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines*, *28*(1), 191–235. https://doi.org/10.1007/s11023-018-9459-4

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* MIT Press. https://doi.org/10.7551/mitpress/6730.001.0001

Wiese, W. (2017). What are the contents of representations in predictive processing? *Phenomenology and the Cognitive Sciences*, *16*(4), 715–736. https://doi.org/10.1007/s11097-016-9472-0

Williams, D. (2017). Predictive processing and the representation wars. *Minds and Machines*, *28*(1), 141–172. https://doi.org/10.1007/s11023-017-9441-6

Williams, D. (2018). Pragmatism and the predictive mind. *Phenomenology and the Cognitive Sciences*, *17*(5), 835–859. https://doi.org/10.1007/s11097-017-9556-5

Williams, D., & Colling, L. (2018). From symbols to icons: The return of resemblance in the cognitive neuroscience revolution. *Synthese*, *195*, 1941–1967. https://doi.org/10.1007/s11229-017-1578-6

Wittgenstein, L. (1953). *The philosophical investigations*. Blackwell.

Zahnoun, F. (2021). The socio-normative nature of representation. *Adaptive Behavior*, *29*(4), 417–429. https://doi.org/10.1177/1059712320922364

Open Access