Synthese (2024) 203:7
https://doi.org/10.1007/s11229-023-04418-6

ORIGINAL RESEARCH

Neural representations unobserved—or: a dilemma for the cognitive neuroscience revolution

Marco Facchin[1]

Received: 21 April 2023 / Accepted: 4 November 2023 / Published online: 20 December 2023

Abstract
Neural structural representations are cerebral map- or model-like structures that structurally resemble what they represent. These representations are absolutely central to the "cognitive neuroscience revolution", as they are the only type of representation compatible with the revolutionaries' mechanistic commitments. Crucially, however, these very same commitments entail that structural representations can be observed in the swirl of neuronal activity. Here, I argue that no structural representations have been observed being present in our neuronal activity, no matter the spatiotemporal scale of observation. My argument begins by introducing the "cognitive neuroscience revolution" (Sect. 1) and sketching a prominent, widely adopted account of structural representations (Sect. 2). Then, I will consult various reports that describe our neuronal activity at various spatiotemporal scales, arguing that none of them reports the presence of structural representations (Sect. 3). After having deflected certain intuitive objections to my analysis (Sect. 4), I will conclude that, in the absence of neural structural representations, representationalism and mechanism can't go together, and so the "cognitive neuroscience revolution" is forced to abandon one of its commitments (Sect. 5).

Keywords Structural representations · Eliminativism · Neuroscience · Mechanistic explanation · Neurocognitive revolution

1 Introduction: neural structural representations and the cognitive neuroscience revolution

Representations remain as central to cognitive science as elusive to our understanding (Villaroja, 2017; Favela & Machery, 2023). Philosophers invested in the "cognitive neuroscience revolution" (Boone & Piccinini, 2016), however, argue that cognitive

✉ Marco Facchin
marco.facchin.marco.facchin@gmail.com; mfacchin@uantwerpen.be

[1] Centre for Philosophical Psychology, Antwerp University, Prinstraat 13, 2000 Antwerp, Belgium

*neuro*science[1] operates upon a stable concept of *neural* representation. In their view, cognitive neuroscience depicts neural representations as inner maps or models that represent their targets by resembling them in a particular, *structural* way. Call these representations neural structural representations—NSRs for short (see Gładziejewski, 2015, 2016; Gładziejewski & Miłkowski, 2017; Williams, 2017; Williams & Colling, 2017; Wiese, 2016, 2017; Morgan & Piccinini, 2018; Piccinini, 2020a, 2020b, 2022).

Prima Facie, contemporary cognitive neuroscience relies heavily on NSRs. The spatial navigational skills of rats are explained by appealing to a cognitive *map* hosted in the hippocampus (cf. O'Keefe & Nadel, 1978; Moser et al. 2008). Motor control is accounted for in terms of various *models* computing and controlling motor trajectories (Mcnamee & Wolpert, 2019; Pickering & Clark, 2014), which might underpin social cognition (Haruno et al., 2003). The "mirror" property of many neurons is increasingly interpreted in terms of inner *models* allowing to simulate actions (Csibra, 2008; Kilner et al., 2007) and emotions (Rizzolatti & Sinigaglia, 2023). Popular neurocomputational frameworks such as predictive processing cast all brain functions as operations on complex, multifaceted statistical *models* of the environment (cf. Buckley et al., 2017).[2] More generally, the idea that inner *models* are the only way in which an agent can make sense and control the flux of input the environment bombards the agent with is gaining momentum (Brette, 2019; Seth, 2015). The cognitive centrality of inner *models* is further confirmed by a host of neurorobotic experiments (Tani, 2007, 2016) and neurocomputational models (cf. Ha & Schmidhuber, 2018a, 2018b; Poldrack, 2020). And so, whilst such map- and model-like structures are in no way the *only* type of representational structure cognitive neuroscientist invoke (cf. Backer et al., 2022; Barack & Krakauer, 2021; Frisby et al., 2023), it is undeniable that they do play a large explanatory role in contemporary cognitive science.

When it comes to the "cognitive neuroscience revolution", however, NSRs are not "just" important. They are *central*. For, supporters of the "cognitive neuroscience revolution" claim that cognitive neuroscience is deeply committed to a *mechanistic* explanatory strategy (see Gładziejewski, 2015; Boone & Piccinini, 2016; Williams & Colling, 2017; Piccinini, 2020a).[3] On this view, to explain cognitive capacities (and their behavioral manifestations) is to identify and describe the physical mechanism responsible of them; that is, to identify and describe a set of organized physical components whose causal interaction constitutes the cognitive capacity in question (or causes the relevant behavioral manifestation; see Craver, 2007; Bechtel, 2008). Crucially, mechanistic explanations are (at least partially) ontic explanations. Their explanantia are not (only) statements concerning mechanisms, but also the *actual* mechanisms (cf. Craver, 2007, p. 27; Illari, 2013).

Clearly, if one accepts this view of explanation in regards to cognitive (neuro)science, it follows that the representations invoked in cognitive (neuro)scientific explanations must be real and literal components of our

[1] Here, "cognitive neuroscience" and "cognitive science" will refer only to mainstream approaches—that is, representational and computational—in the respective disciplines. For non-mainstream alternatives, see (Anderson 2014; Bruineberg & Rietveld 2019; Chemero 2009; Kelso 1995; Van der Weel et al., 2022).

[2] Predictive Processing also admits non-representational interpretations which (sadly) remained quite marginal (see Downey 2018; Facchin 2021a).

[3] But see (Silberstein & Chemero 2013; Silberstein 2021) for a diverging opinion.

(neuro)cognitive mechanisms, whose content must literally and really be causally efficacious within the mechanisms's inner functional economy—which is *prima facie* highly problematic. For, it is quite natural to think that representational contents are causally inert. All the heavy causal lifting seems done by the representational *vehicles*—the physical structures "doing" the representing by "carrying" the contents around—rather than the contents themselves (cf. Egan, 2020). So, aren't mechanistic explanations just *incompatible* with content-based, representational explanations?

Proponents of the "cognitive neuroscience revolution" appeal to NSRs to provide a negative answer. For, structural representations are underpinned by representational vehicles whose physical shape is not just casually potent, but also *semantically relevant*. This is because the physical shape of the vehicles, and the particular way in which they resemble their targets, determines what these vehicles represent. In the case of NSRs, then, semantic content and vehicular shape are at least largely overlapping, if not the exact same thing (Lee, 2019; Piccinini, 2022; Williams & Colling, 2017). Semantic contents are thus able to play an active causal role within our neurocognitive mechanisms, and are thus able to play a genuine explanatory role in mechanistic explanations (cf. O'Brien, 2015).

NSRs are thus cast as genuinely representational components of neurocognitive mechanisms, allowing representational and mechanistic explanations to mesh. This view, I now want to highlight, has an important implication: if NSRs are bona fide components of neurocognitive mechanisms, then they must be observable and manipulable as any other component of said mechanisms. Proponents of the neurocognitive revolution agree—either implicitly (see Williams, 2017) or explicitly (Piccinini, 2020a, Thompson and Piccinini 2018). Hence, at least insofar NSRs are concerned, we can circumvent the seemingly never-ending debate concerning the reality of internal representations (cf. Anderson & Champion, 2022; Hutto & Myin, 2013; Ramsey, 2007; Segundo-Ortin & Hutto, 2021). To determine whether NSRs are real, one just needs to peek inside the neurocognitive system and see whether NSRs—or, more accurately, NSRs-supporting vehicles[4]—can be found (cf. Bechtel, 2008, 2014; Facchin, 2021a; Piccinini, 2020a; Thomson & Piccinini, 2018). For simplicity, let me refer to NSRs supporting vehicles as NSRVs.

The aim of this paper is to take one such peek. As its title suggests, I argue that no NSRVs can be observed. My analysis unfolds as follows. (Sect. 2) introduces a widely accepted account of structural representations, focusing on the constraints it places on representational vehicles. (Sect. 3) considers whether neuronal vehicles satisfy these constraints, focusing in particular on activations of individual neurons (Sect. 3.1), neural maps (Sect. 3.2) and activation spaces (Sect. 3.3). In all these cases, I conclude that the relevant vehicles do not satisfy the constraints introduced in (Sect. 2), and so that they can't be NSRVs. (Sect. 4) anticipates some objections. (Sect. 5) considers the implications of my verdict for the cognitive neuroscience revolution, concluding the paper.

[4] This caveat is actually important: NSRs proper are *relations* between neural vehicles and their targets, so they can't be observed *just* by observing neural goings on. At best, then, observing neural goings lets us see one *relatum*, that is, the relevant representational vehicles (the NSRV).

2 A standard account of (neural) structural representations

Informally described, structural representations are model- or map- like structures which represent their targets (i.e. what the representation is "aimed at") by being structurally similar to them. Cartographic maps are paradigmatic examples of structural representations, for they represent a terrain *by replicating* the terrain spatial structure with their own spatial structure: if location *a* is west of location *b*, then the map will display the point standing for *a* left of the point standing for *b*. Can this intuitive, but imprecise, idea of a structural representation be made more rigorous?

Paweł Gładziejewski (2015, 2016) offers a nowadays standardly accepted philosophical analysis of structural representations[5]:

Within a system **S**, a vehicle **V** is the vehicle of a structural representation of a target **T** if and only if:

(1) **Structural similarity: V** is structurally similar to **T**; &
(2) **Action Guidance:** The structural similarity in **(1)** allows **V** to guide **S**'s action in regards to **T**; &
(3) **Decouplability: (2)** can obtain even when **V** is decoupled from **T**; &
(4) **Error Detection: S** can detect the representational errors of **V**

There is much to say about **(1)**-**(4)**, both as individually and as a whole. One first important thing to notice is that they all concern structural representations *in general*—they're not specific to NSRs. This is a good thing, as it allows me to explain **(1)-(4)** in terms of structural representations everyone is familiar with, such as maps. The step from structural representation in general to NSRs can then be easily made by placing an appropriate restriction on the physical medium realizing the vehicles: vehicles must be realized by neurons—or, more precisely, by patterns of neuronal activities.

Secondly, **(1)**-**(4)** all concern the *vehicle* of a structural representation. Consider, for example, the physical support underpinning a cartographic map. It is that support—that is, the representational vehicle—that **(1)** is structurally similar to the mapped terrain, **(2)** is used to guide our actions (e.g. in traversing said terrain), **(3)** can guide our actions when we're decoupled from that terrain (e.g. allowing us to plan the way ahead), and **(4)** whose usage allows us to detect its eventual representational errors (e.g. by noticing that it leads us systematically astray). So, **(1)**-**(4)** specify the relevant vehicular features underpinning structural representations. Notice also that, since **(1)**-**(4)** are imposed in conjunction, the vehicles underpinning structural representations must satisfy all of them. I will now examine each condition in turn, focusing in particular on **(1)** and **(2)**, as they will be extremely important throughout the entirety of (Sect. 3).

Condition **(1)** requires the representational vehicle **V** to be structurally similar to the represented target **T**. Note that the relevant similarity relation holds between a *single* vehicle and its target, rather than *a number* of vehicles and a set of targets. To unpack the relevant similarity relation, like Gładziejewski, I chose a very liberal unpacking,

[5] Through, as a reviewer noticed, this is not the *only* possible understanding of structural representations. See the *Appendix a*t the end of the paper. Still, Gładziejewski's account remains the one most typically referred to in the cognitive neuroscience revolution.

which makes **(1)** *easier* to satisfy—and so, NSRVs easier to spot.[6] Thus, this is the relevant charitable interpretation of NSRs in the present context. On my view:

> **V** is structurally similar to **T** if and only if:
> **(a)** There is a one-to-one mapping from some vehicle constituents ($V_A \ldots V_N$) of **V** to some target constituents ($T_A \ldots T_N$) of **T**; &
> **(b)** There is one relation $\Re$ holding among the vehicle constituents of **V** and one relation Γ holding among the target constituents of **T** such that, for all the vehicle constituents satisfying **(a)**: $\Re(V_A,V_B) \rightarrow \Gamma(T_A,T_B)$. (cf. O'Brien and Opie 2004).

(a) imposes a *one-to-one* mapping from some relevant physical bits and pieces of the vehicle **V** (i.e. vehicle constituents) to some bits and pieces of the target **T** (i.e. the target constituents). I won't pose any restriction on what may count as a vehicle constituent—everything may be vehicle constituent, provided that it is a material constituent of a vehicle. For the sake of simplicity, however, I won't consider here arbitrary, or "unnatural" way of carving up vehicles: whilst "unnatural" mappings alway allow to find a structural similarity (cf. McLendon, 1955), it is very doubtful our neurocognitive systems *care* about them—they won't be, as Shea (2018) usefully puts it, *exploitable* by our neurocognitive system. Also, again for the sake of simplicity, I'll always assume that the mapping in **(a)** is "subscript preserving": V_A maps onto T_A, V_B maps onto T_B, … and V_N maps onto T_N.

(b) forces **V** and **T** to share the *same inner relational structure*: if a relevant relation holds between V_A and V_B and they satisfy **(a)**, then a relevant relation holds between T_A and T_B. Notice that **(b)** mentions *one* relation in **V** and *one* in **T**. So, in order for **(b)** to obtain the relations preserved by the mapping in **(a)** needs to be constant on both sides of the mapping. So, if $\Re(V_A,V_B) \rightarrow \Gamma(T_A,T_B)$ but $\Re(V_C,V_D) \rightarrow \backslash(T_C,T_D)$ rather than $\Gamma(T_C,T_D)$, then **(b)** fails to obtain and **V** and **T** are not structurally similar. To understand the point intuitively, imagine a map representing the distance between some cities in a region in terms of distances between them, *and also* the distances between other cities in the same region only in terms of the *colors* used to represent the cities (e.g. cities represented in *darker* colors are further apart than cities represented in *lighter* colors). Such a map *would not* count as a structural representation according to Gładziejewski's analysis (and it would also be *really* hard to use).[7]

Crucially, conditions **(a)** and **(b)** determine the relevant semantic properties of structural representations. They determine what a vehicle **V** represents.[8] In structural representations, V_A *represents* T_A, and the fact that $\Re(V_A,V_B)$ *represents that* $\Gamma(T_A,T_B)$ (e.g. Shea, 2018). Again, this point can be made intuitively clear by looking at ordinary cartographic maps. Imagine a map of the Alps. The biggest triangle shaped

[6] As Kohar (2023) has persuasively argued, this is also the only *relevant* unpacking of the structural similarity.

[7] As an additional point, notice that **(b)** allows for $\Re=\Re$. So, the two relations *can* be identical. And that is exactly what happens with regular cartographic maps, in which *spatial* relations are involved on both sides of the mapping.

[8] At least partially. Other factors may be relevant in determining the content of **V**. For example, Shea (2018) calls upon teleological factors, whereas Piccinini (2020a; 2022) calls upon teleo-informational factors and factors concerning the embodiment and embeddedness of cognitive systems.

figure on the map *represents* the biggest mountain of the Alps (Mont Blanc), and the fact that such a big square is placed above a big green area *represents* the fact that Mont Blanc is north of the Italian plains.Thus **(a)** and **(b)**—that is, **(1)**—are the reasons why the physical shape of the representational vehicles of structural representations are imbued with their semantic properties (Williams & Colling, 2017).

Notice how **(1)** entails that structural representations have a specific form of semantic transparency. Since the mapping in **(a)** is one-to-one and **(b)** operates only on *one* relation for **V** and one relation for **T**, then it is always possible to interpret all the "$\Re(V_X,V_Y)$" univocally and transparently: $\Re(V_A,V_B)$ can only represent $\Gamma(T_A,T_B)$. Notice that since structural representations are transparent, their content is neither disjunctive nor indeterminate: $\Re(V_A,V_B)$ *represents that*—and only that—$\Gamma(T_A,T_B)$. Were it to represent something disjunctive or indeterminate—say, something like $\{\Gamma(T_A,T_B)$ *or* $\Gamma(T_@,T_B)\}$ or $\{\Gamma(T_A,T_B)$ *or* $\searrow(T_A,T_B)\}$—then either **(a)** or **(b)** would fail to obtain, and with it **(1)** would fail to obtain too.

Condition **(2)** is satisfied when the structural similarity in **(1)** guides the actions of a system **S** that are "aimed at" **T**. When this happens, **S**'s odds of success are sensitive to the *quality* of the similarity holding between **V** and **T** (see Shea, 2018, p.142). The more **V** structurally resembles **T**, the higher **S**'s odds of non-accidental success; and, the lower the quality of the resemblance, the lower **S**'s odds. *Ceteris paribus*,[9] the better the map resembles the terrain, the more one is able to traverse it. The worse their resemblance, the more one is likely to get lost.

Notice that satisfying **(2)** entails that content is causally potent. For, intervention on the structural similarity in **(1)** *just are* interventions on what **V** represents—that is, its *contents*. But, as seen above, these interventions also modify the agent's odds of success: the better the similarity, the better the agent's odds. This is enough to make **V**'s content causally potent under an interventionist notion of causality (Gładziejewski & Miłkowski, 2017): changes in **V**'s contents *cause* an agent to be more likely to non-accidentally succeed or fail.

Here, I wish to highlight two ways in which the structural similarity between **V** and **T** can be worsened—and so, two ways to non-accidentally decrease an agent's odds of success. First, the similarity can be worsened because single vehicle constituents of **V** map onto *many* target constituents of **T**. This is one way to violate **(a)**. I will call it an *(a)-violation*. Secondly, the similarity between **V** and **T** may be degraded because two constituents of **V** display the corresponding constituent of **T** as being in a relation that does not in fact hold. This is one way of violating **(b)**—and I will call it a *(b)-violation*. Resorting to the map example may help clarify both cases. When an (a)-violation occurs, one bit of the map "stands for" multiple bits of the terrain—like a dot on a map representing *both* Paris and Rome. When a (b)-violation occurs, the map inaccurately displays the terrain by displaying certain parts of it being in a relation that does not in fact hold between them—like a map displaying Rome north of Paris. There are of course *further* ways in which the structural similarity between **V** and **T** may be worsened: **(a)** and **(b)** can be violated in many other ways. But my arguments won't hinge on *these* violations, so I won't discuss them.

[9] This *Ceteris paribus* clause is meant to exclude cases in which excessive degrees of similarity stand in the way of representational usage, as in the case of an hypothetical map in 1:1 scale.

Point **(3)** mentions decouplability. Decouplability is an essential feature of all representations, which captures the idea that representations represent their target even when their target is not causally affecting them or the agents relying on them (cf. Orlandi, 2020). A map can be used even when the mapped terrain is not causally interacting with the map or its user: for example, a map of Tokyo represents Tokyo even if it, and its user, are located in Buenos Aires. Minimally, then, decouplability can be unpacked as follows: **V** is decoupled from **T** when **T** is not causally influencing **V**—for example, by causing its tokening (cf. Gładziejewski, 2015, 2016). Notice, however, that **(3)** requires something *more* than decouplability thus spelled: it requires decouplable representations to still play the action guiding role specified by **(2)** when decoupled. So, for a map of Tokyo to fully satisfy **(3)** it is not enough that it continues to represent Tokyo while located in Buenos Aires. It must also perform its action guiding duties while in Buenos Aires—for example, by allowing the map user to plan her trip to Tokyo in a way such that the plan's odds of non-accidental success depend on the degree of similarity holding between the map and Tokyo.

Lastly, **(4)** is entailed by **(2)**[10]: if **V** guides **S**'s actions in regards to **T** as required by **(2)**, then the degree of similarity between **V** and **T** is reflected in **S**'s odds of success. Hence non-accidental behavioral successes can act as reliable (through defeasible) indicators of representational accuracy: pragmatic successes indicate representational successes, and pragmatic failures indicate representational failures—thereby allowing the detection of representational errors. (cf. Gładziejewski, 2015, 2016, see also Bielecka & Miłkowski, 2020 for further elaboration).

Summing up: structural representations are representational vehicles **(1)** structurally similar to a target, **(2)** whose structural similarity guides an agent's action aimed at that target, **(3)** that can do so even when decoupled from their target and **(4)** that allow their user to determine their representational accuracy via the success-rate of the actions they guide. NSRs are just structural representations realized in the neural medium. Thus, if they are present, we should be able to observe NSRVs: neural vehicles satisfying **(1)**-**(4)**.

But, does our neuronal activity *really* realize such vehicles? I think the existing neuroscientific data motivate a negative answer.

3 Are bona fide neural vehicles vehicles of neural structural representations?

To determine whether neural vehicles satisfy **(1)-(4)**, one must first determine what neural vehicles are. Here, I take neural vehicles to be neural activity, which I will consider at various nested scales of spatiotemporal organization. Now, neural activity *need not* be triggered by any external stimulus or task: its origin may be endogenous, and entirely determined by the inner dynamics of the nervous system. The activity of the "resting state" or "default mode" network may be a good example of such purely endogenous neural activity (cf. Raichle, 2015). Understanding such bouts of intrinsic

[10] At least, in sufficiently complex systems: we surely could design a robot whose central control system allows the tokening of states satisfying **(1)-(3)** but not **(4)**. However, since the paper focuses on brains (and brains are arguably *sufficiently complex*) I will take **(4)** to be entailed by **(2)**.

activity is surely crucial to fully understand how the brain works. Yet, in this paper, I will largely ignore them. As the majority of neuroscientists (eg. Friston, 2005; Mesulam, 2008; Villaroja, 2017; Backer et al., 2022; Frisby et al., 2023) and philosophers of mind/cognitive (neuro)science—defenders of the "cognitive neuroscience revolution" included (e.g. Piccinini, 2020a; Thomson & Piccinini, 2018)—I will here be concerned mostly with neuronal *responses* to external stimuli. I will analyze them at three distinct spatiotemporal levels: the level of individual neuronal responses (Sect. 3.1), the level of neural maps (Sect. 3.2), and the level of entire activation spaces (Sect. 3.3). In all these cases, I claim that they do not, and, indeed, *cannot*, satisfy **(1)-(4)**.

Why focus on neuronal responses? The reason is purely *methodological*. Pretty much every form of representationalism has to assume that, when a system **S** deals with a target **T**, **S** has to tokenize a vehicle **V** which represents **T**—else, a representational explanation of the phenomenon would not be warranted. So, observing how **S** responds to **T** is a *good* way to observe what vehicle **V** (representing **T**) **S** tokenizes, in a way that makes it easy to check whether **V** is a NSRV. Observing intrinsic bouts of **S**'s activity, on the other hand, is a *less reliable* observational procedure. For, we can't exclude that at least some of these bouts of activity are *not* representational vehicles, and we currently lack a way to tell these two apart. Moreover, we presently lack a way to connect "endogenously" tokenized vehicles to their targets, in a way that clearly prevents us from determining whether they qualify as NSRV.[11]

That being said, I can't help but admit that neuronal responses are *not* the only representational vehicles populating our brain. Indeed, alongside bouts of "intrinsic" neural activity, other neuronal activities and structures are labeled as neuronal vehicles. I will consider some of these structures in (Sect. 3.4), arguing that these too fail to qualify as NSRV.

3.1 Individual neuronal responses are not vehicles of neural structural representations

Individual neurons respond to stimuli selectively: different stimuli elicit different responses. Typically, neurons have one *preferred stimulus*, which elicit the strongest response. Preferred stimuli vary depending from neuron to neuron, reflecting their specialized functional roles. For example, neurons in the primary visual cortices respond to simple visual stimuli like oriented bars (cf. Hubel & Wiesel, 1968). Neurons in hierarchically higher layers of the visual cortex respond to more complex stimuli—for example, neurons in area MT respond to movement directions (cf. De Angelis & Newsome, 1999). Neurons further away from sensory areas respond to even more complex stimuli (or features thereof): the parietal cortex houses neurons responding to specific quantities (Nieder et al. 2006), the inferior premotor areas of the frontal cortex house neurons that respond to specific actions (Kohler et al., 2002) and, apparently, there are even neurons in the inferior temporal cortex preferring specific individuals

[11] Notice that the point here is *exclusively* methodological. It should not be confused with the endorsement of an "indicator" view of representation, according to which neural activity represents what it causally sensitive to/ correlates with. On the relationship between structural representations and indicators, see references given in (**§3.1**).

(Quiroga et al., 2005). Thus, individual neurons have preferred stimuli of different sorts, which they are often said to *represent*. But are these representations NSRs? Are they underpinned by NSRVs?

It is a bit hard to provide a direct answer to these questions. Sure, NSRVs should be observable and manipulable as any other component of a mechanism—but this time it is a bit unclear what we should be looking at (or thinking with) *exactly*. For, "individual neuronal response" can be read in at least three different ways: (i) as designating individual spikes (i.e. single neuronal discharges), (ii) as designating spike trains (i.e. sequence of discharges) and (iii) as designating a neuron's firing rate compared to a baseline. Options (i)-(iii) all pick up a *bona fide* representational vehicle supporting a specific representational scheme (e.g. Brette, 2015; Dayan & Abbott, 2005). Thus, the claim that individual neuronal responses are NSRVs can be read in at least three different ways. As a consequence, it is not immediately clear what sort of observations and manipulations would support it.[12]

Now, whilst interpretations (i)-(iii) are all possible, I want to suggest that they all face certain important challenges, whose collective weight seems enough to reject the idea that individual neuronal responses may qualify asNSRVs under *any* interpretation.

First, it is very hard to see how an individual neuronal response could structurally resemble its target—be it an oriented bar or an individual person. This is because it is very hard to see how the vehicle (i.e. the individual response, however interpreted) could be non-arbitrarily decomposed into vehicle constituents as requested by **(a)**. It is not at all clear what could count as a vehicle constituent of a single neuronal response: a "part" of a spike, an individual spike (or sequence of spikes) in a spike train, part of the voltage emitted, a fraction of the firing rate, part of the neurotransmitters emitted, or something else entirely? All these options pick up certain bona fide parts of a single neuronal response. Yet, there seems to be no privileged way to choose between them (cf. Maley, 2023): the choice of vehicle constituents seems entirely arbitrary. This is a serious problem when it comes to satisfying **(1)**. Of course, I don't want to deny that we may *discover* that there are functionally relevant, non-arbitrary ways to decompose individual neuronal responses. But we've not discovered them yet. So, even supposing that one such partition exists (which is something my dialectical adversaries should argue for!) we've not yet observed the relevant NSRVs, for we simply do not know what that partition is. Moreover, even if a privileged, non-arbitrary way to identify vehicle-constituents in individual neuronal responses were to be found, we still would have to specify what sort of relevant *relation* holding amongst the vehicle-constituents as specified by **(b)**. A task as daunting as the previous one.

Secondly but not least importantly, such tasks are not just daunting. They are also entirely unmotivated—at least insofar the explanatory practices of present day cognitive neuroscience go. For, whilst contemporary cognitive neuroscientists typically assume that individual neuronal responses represent individual targets, they do *not* claim that specific *parts* of neuronal responses represent specific *parts* of the target, nor do they claim that relations between parts of neuronal responses represent relations

[12] Notice that the claims that neuronal maps and activations spaces are vehicles of NSRs are not similarly ambiguous: both claims express a form of population coding, which is a special case of *rate* coding. No interpretation of these claims in terms of single spike trains (or single spikes) is possible.

between parts of the target. But that's exactly the way in which structural representations represent. Moreover, I suspect that claims such as "The first spike of the spike train represents the leftmost bit of the oriented bar" or "the fact that spike V_A preceded spike V_B represents the fact that a part T_A of the oriented bar is left of a part T_B of the same bar" would be considered not just unjustified, but entirely *exotic* by the majority of cognitive neuroscientists. So exotic, indeed, to be a *bona fide reductio* of the idea that individual neuronal responses are NSRVs.[13]

Summing up: the claim that individual responses are NSRvs is hard to "cash out", it yields extremely exotic conclusions and it is entirely unjustified by the current practice of cognitive neuroscience. Individual neuronal responses are in fact typically described as "indicator" or "detector" representations (cf. Backer et al., 2022; Gładziejewski & Miłkowski, 2017; Ramsey, 2003; Williams & Colling, 2017).[14] On this view, the firing of a neuron does not provide an inner model of a target which replicates the target's inner structure. Rather, the firing of a neuron simply signals the presence of the target at the time of firing. So, the actual practice of cognitive neuroscience—that is, the observations and manipulations that cognitive scientists actually carry out—does not suggest or motivate the claim that individual neuronal responses are NSRVs. If anything, individual neuronal responses are said to be the *vehicle constituents* of individual structural representations (cf. Gładziejewski & Miłkowski, 2017; Williams & Colling, 2017)—a view whose two different popular incarnations will be discussed in (Sects. 3.2 and 3.3).

Before moving on, however, I need to face an increasingly popular line of argument, which casts indicators as a special case of structural representation (Morgan, 2014; Facchin, 2021b, Nirshberg & Shapiro 2021, Nirshberg, 2023).[15] Clearly, this view challenges my analysis: if indicators are structural representations, then indicator states qualify as NSRV. And so, if individual neuronal responses are indicator states, to observe them *just is* to observe a NSRV.

Why, however, should one think that indicators are a special case of structural representation? The answer seems to be the following. Consider an indicator such as a familiar mercury thermometer. The height of the bar grows proportionally to the temperature: the *hotter* the environment, the *higher* the bar. Such an observation seems to generalize to all indicators: there is always some indicator-specific relation between indicator states and indicated states. In a hygrometer, the *higher* the humidity, the *longer* the indicating hair gets; the floater of a fuel gauge gets *lower* as the tank gets *emptier*, and so forth. So, for every pair of indicator states V_A and V_B there is a relation such that the corresponding indicated states T_A and T_B are in a relation. In other words, $\Re(V_A,V_B) \rightarrow \Gamma(T_A,T_B)$ holds, and the indicator and the indicated target are thus structurally similar. This is enough to satisfy **(1)**. And various ingenious

[13] One could still argue that individual neuronal responses represent what they represent *because* they are part of a larger structural representation. Notice, however, that, in such a case, individual neuronal responses would not be NSRVs, but only *vehicle constituents* of a larger NSRV. At any rate, **§§ 3.2-3.4** will consider putatively larger vehicles, concluding that they don't qualify as NSRVs either.

[14] Piccinini (2020a) might, under a certain reading, be an exception—but he really seems more concerned with *populations* of neurons rather than individual neurons. I will thus deal with his view in (**§3.2**).

[15] See also (Gładziejewski & Miłkowski 2017; Lee & Calder 2023) for other attempts to resist this view.

arguments can take care of **(2)**—**(4)**. Or so, at least, many philosophers have argued (see, in particular, Facchin, 2021b).

Should we thus accept that indicators are structural representations? I'm inclined towards a negative answer—but I won't defend it here, as determining whether indicators *really* are structural representations is clearly beyond the scope of this paper.[16] What I will notice, however, is that even if the line of reasoning sketched above were successful, it *would not* show that individual neuronal responses satisfy **(1)**. To see why, consider *how* that pattern of reasoning may be applied to individual neuronal responses. Imagine to observe a neuron that indicates an oriented bar: the neuron fires *most vigorously* as the orientation of the bar triggering it *most closely approximates* the preferred orientation, hence $\wr(V_A,V_B) \rightarrow \Gamma(T_A,T_B)$. But what are V_A and V_B in this example? The natural answer is that they are the *individual neuronal responses* to stimuli T_A and T_B. When the neuron "saw" the stimulus T_A (which approximates the preferred orientation better than T_B) then it responded entering in a state V_A (which is much more active than V_B). So, *according to the very same pattern of reasoning that should demonstrate that indicators are structural representations*, individual neuronal responses fail to qualify as structural representations—at least, given the relevant notion of structural representation discussed in Sect. 2. They are, at best, *constituents* of a structural representation. It follows that observing individual neuronal responses does not amount to observing NSRV.

At this point, however, it is natural to wonder whether **V**—the entire set of indicator states being structurally similar to **T**—qualifies as a NSRV that we have observed. The answer to this question will be provided in (Sect. 3.4). But before providing it, I need to consider another important (family of) candidate NSRV; namely *neural maps*.

3.2 Neural maps

Above, I've argued that individual neuronal responses are not NSRVs. But what about the responses of *multiple* neurons?

Piccinini (2020a), argues at length that various types of cortical maps—including the retinotopic map in the primary visual cortex and the motor and sensory *homunculi*—qualify as NSRVs. Ramsey (2016), Shea (2018) and Gładziejewski and Miłkowski (2017) all claim that certain neurons in the hippocampus of rats are connected in a map-like way, so as to structurally represent the rat's environment.[17] So, many authors suggest that the real NSRVs are responses of *multiple* neurons organized in a map-like way.

These arguments can call upon a wealth of well-known neurophysiological and neuropsychological data. For example, Piccinini (2020a, p. 271) stresses the retinotopic organization of the primary visual cortices (V1), nicely displayed in Fig. 1:

The neurons constituting V1 them are spatially organized so as to replicate (a tweaked version of) the spatial structure of the original visual stimulus (cf. Tootell et al., 1988). If neuron V_A is *left of* neuron V_B, then T_A (i.e. whatever V_A is responding

[16] Through see the *post scriptum* to see one reason underpinning such a negative answer.

[17] See (Bechtel 2008; 2014; Thomson and Piccinini 2018) for a non NSRs-centric representational account of these neural structures.

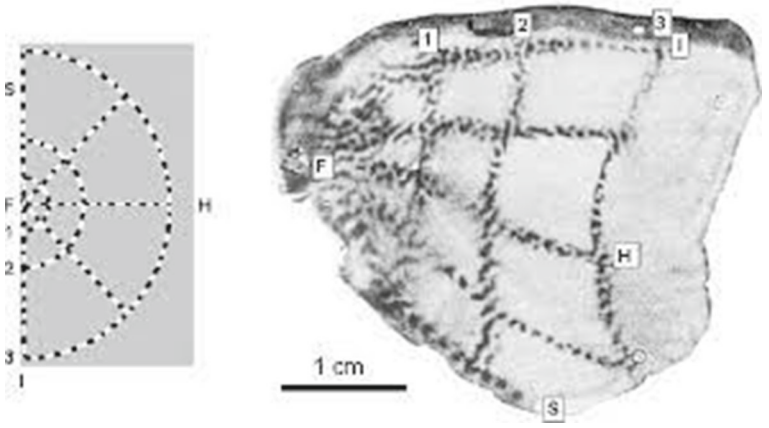

Fig. 1 Cortical topography of V1: the spatial structure of the stimulus (left) is mirrored—in a systematically distorted fashion—by V1 neurons (right). The same topological structure, however, is instantiated in both. Source: Figs. 1 and 2b in (Tootell et al. 1998). Reproduced with permission. Copyright (1988) Society for Neuroscience

to) is left of T_B. This is a clear structural similarity tying together the neural map and its representational target. Further, Piccinini stresses that the columnar organization of V1 contains many "smaller scale" cortical maps representing significant properties:

> "V1 contains multiple fine-grained topographically organized feature maps of such properties embedded in the larger-scale retinotopic representation of space. For instance, those neurons selective for horizontally oriented bars tend to cluster together in cortical columns in V1, and nearby columns contain neurons that are tuned to similar orientations" (Piccinini, 2020a, p. 272).

So, if column V_A is close to column V_B, then T_A is similar to T_B. Similar "smaller scale" maps are found in many neural areas. For example the neurons area MT (a further visual area particularly sensitive to movement) are arranged so as to compose a "movement map". Neurons that prefer similar direction of motion cluster into columns, and columns are spatially organized so that spatially close columns prefer similar movements (cf. De Angelis & Newsome, 1999). The closer two columns (or two neurons) are, the more similar the velocities they respond: if V_A is close to V_B. then T_A is similar to T_B. More intuitively strikingly still, there are the cortical "homunculi" and "simunculi" drawn by Penfied and Woolsey (cf. Penfield and Brodley 1937; Woolsey et al., 1952). It's hard to look at them without noticing how nicely the spatial organization of these neurons "recapitulates" the spatial organization of bodily parts—for one example, see Fig. 2.

Notice how easily the relevant structural similarity can be *seen* in Fig. 1 and Fig. 2. Isn't it simply obvious that these structures *are* structurally similar to their target, in a way that clearly satisfies **(1)**?

Whilst these structures obviously *seem* structurally similar to their targets, it is not at all obvious that they *are*—or so I will later argue. But before doing so, I wish to notice that even if such similarities were present, the fact that their presence is obvious *to us*

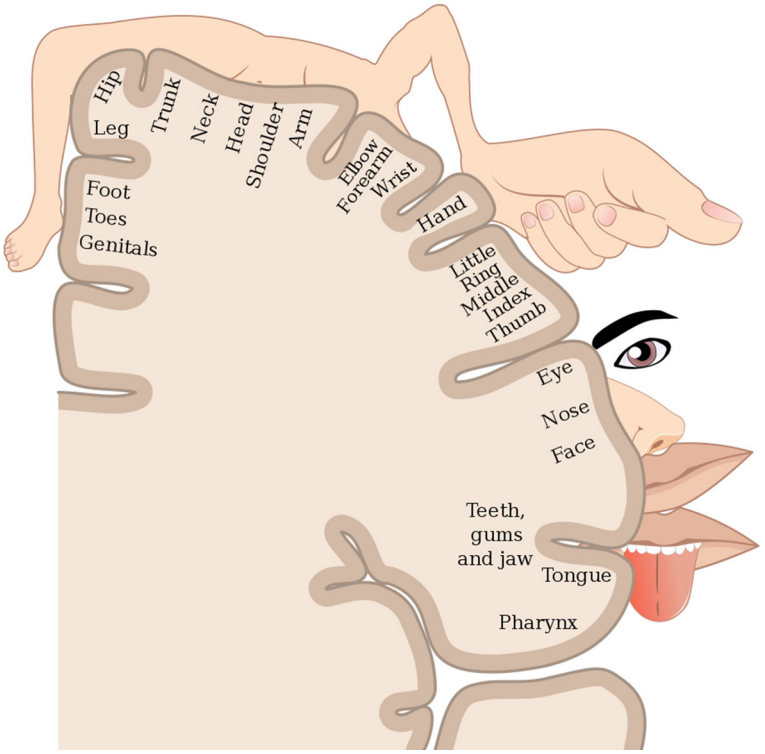

Fig. 2 The sensory homunculus. Note how the spatial relations between the cortical areas "mirror" the spatial relations between the represented body parts
Source: Wikimedia. This file is reproduced, without modification, from https://commons.wikimedia.org/wiki/File:Sensory_Homunculus-en.svg in accordance with its Creative Commons License. Original creator: OpenStax College (http://cnx.org/content/col11496/1.6/). Modified by Wikimedia users Popadius and Preoptic

does not entail that our neurocognitive mechanisms "see" the similarity too—indeed, it seems that our neurocognitive mechanisms are blind such similarities in the execution of their tasks. And, for this reason, these similarities fail to satisfy **(2)**. Consider, for example, the somatotopic organization of the cortical homunculus—and the structural similarity it underpins. Does the somatotopic *spatial arrangement* of these neurons guide our actions as required by **(2)**? *Prima facie*, the answer is negative. Imaginary interventions that modify *only* the somatotopic organization of the homunculi (i.e. the relative spatial locations of the neurons constituting it) do not seem to have any effect on our behavior. After all, if they modify *only* the somatotopic organization of these neurons, they leave intact their input–output profile and mutual connections, allowing the homunculus they constitute to contribute to an agent's behavior *in the same way* in which a somatotopically non-modified homunculus would. Changes in the somatotopicity of homunculi—and the structural similarity they underpin—do neither increase nor decrease the agent's chance of success. So, **(2)** fails to obtain.

One could object that similar though experiments are ill-suited to determine whether **(2)** obtains or not. Looking at some *real* experiments, however, yields the same verdict. Consider, for example, the data collected by Hartmann et al. (2016).[18] Simplifying to the extreme, they equipped rats with prosthesis enabling them to perceive and respond to infrared lights. The prosthesis were "caps" of infrared sensors (allowing for a 360° panoramic infrared vision) that communicated with the rat's "sensory homunculus" (i.e. their primary somatosensory cortices). Crucially, they could do so in a way that either respected or flouted (to various degree) the homunculus's somatotopic organization—e.g. the front-facing infrared sensor could be connected with the head of the rat's homunculus (respecting somatotopicity) or with its rear or side (flouting somatotopicity). Now, Hartmann and colleagues report that all rats managed to achieve a high success rate in the experimental task (infrared light discrimination), regardless of the degree of somatotopicity of their prosthesis. Sure, the better the somatotopicity, the *faster* behavioral success came. But, eventually, even rats equipped with "non-somatotopic" prosthesis were eventually able to perform at the level of rats equipped with "somatotopic" prosthesis. This clearly violates **(2)**, according to which the degree of somatotopicity should be reflected in higher or lower odds of behavioral success.

Now, one could object that these data are less clear cut that I'm making them appear—after all, rats with "non-somatotopic" prosthesis learned how to face the experimental task more slowly than rats with "somatotopic" prosthesis, and this might be counted as one way in which degrees of somatotopicity influence the agent's odds of non-accidental success during the learning phase. I'm not persuaded that this is the case (why should the degree of somatotopicity matter *only* during the learning phase of a task?)—but even if it were the case, other experimental data on *homunculi* can be marshaled to support my conclusion. For example, Chakrabarty and Martin (2000) have found that, during postnatal development of the primary motor cortex (i.e. the motor homunculus) the number of sites representing more than one limb *increases*. This suggests that such "multi target" sites are needed to effectively control movements—something that improves during postnatal development. And yet, "multi target" neurons clearly degrade the structural similarity in **(1)**, as they are a case of an (a)-violation (cf Sect. 2).[19] So, a *worsening* of the structural similarity correlates with an *increase* of performance, blatantly violating **(2)**. Martin et al (2005) present similar data, suggesting that increases in "multi target" neurons are positively correlated with increases of motor expertise.[20]

The evidence above gestures towards a point that can perhaps be less messily expressed (and generalized beyond homunculi) as follows. The structural similarity of cortical maps is based on certain *spatial* relations holding amongst the map's constituents—that is, spatial relation between neurons. Now, according to a standard

[18] Though it should be noted that the experimental interventions in (Hartmann et al., 2016) are not interventions *only* on somatotopicity, as they always also change the artificial sensors from which neurons receive inputs. Here, I will ignore this complication for the sake of simplicity.

[19] More on this point below.

[20] One could object that motor homunculus is not a good example, because it is not at all clear how the primary motor cortex represents our body and its movements (cf. Piccinini 2020a; Thomson and Piccinini 2018). This, however, is more a problem for the defender of NSRs than for me: how can they claim that the motor homunculus is a NSRV if they do not know what it is structurally similar to?

neuroscientific picture, neurons and neural maps contribute to an agent's behavior in virtue of their information-signaling properties; roughly, their input–output profile. Their input output behavior is determined by a number of features, including a neuron's sensitivity to stimuli, their baseline firing rate, the connections they have with other neurons and the nature of such connections (excitatory or inhibitory) and other features. *Spatial* features, however, do not influence their input–output profile. So, they don't contribute to an agent's behavior. Hence they can be varied ad libitum, creating arbitrarily large (b)-violations, without influencing an agent's behavior and its odds of success. And this, of course, means that they do not play the action guiding role required by **(2)**.

Notice that I'm *not* claiming that the topographic organization of cortical maps does not play any relevant functional role. Not all functional roles of neuronal structures must be representational or cognitive (Haueis, 2018). Perhaps the topographic organization of cortical areas minimizes wiring, speeding up neural signaling (cf. Blauch et al., 2022).[21] Maybe it reduces metabolic costs (cf. Sterling & Laughlin, 2015). Or perhaps it is just a side effect of certain relevant evolutionary or developmental constraints—or maybe it is due to all three, and perhaps even more, factors simultaneously (Cf. Graziano & Aflalo, 2007, p. 239). I'm not denying these (or similar) claims. I'm only denying that the topological organization plays the representational role **(2)** captures. This is entirely compatible with it playing *other* biological—or even cognitive—roles (cf. Graziano, 2011). To deny a car's brakes makes it accelerate is not to say brakes are useless!

One could retort that the argument above is not fully general. In the case of the spatial map in the rat hippocampus, for example, what matters are not the *spatial* relations amongst neurons, but rather the relation of *inducing activation*. If neuron V_A tends to induce the activation of V_B, then T_A is close in space to T_B (cf. Moser et al., 2008). This is a *functional* relation, the changing of which changes the way in which inputs are turned into outputs. Hence **(2)** seems to obtain, and the argument provided above does not apply. And, perhaps, some similar functional relation might similarly rescue the neural maps discussed above. For example, the motor homunculus might not underpin a NSR of our body, but rather a NSR of our action (cf. Graziano, 2016). If that were the case, my focus on somatotopicity might just have distracted from some *other* functionally relevant structural similarity.

Even if that were the case, however, there would still be a significant problem. In general, neurons (including the neurons of cortical maps) do not respond to just *one* stimulus. Sure, they respond most strongly to their preferred stimuli, but it makes sense to say that neurons have *preferred* stimuli only because they respond to many different stimuli. Moreover, the response profile of neurons is typically influenced by *multiple* parameters of a stimulus. For example, MT neurons are not just sensitive to motion direction, but also the retinal position of the stimulus, its size, the speed of motion and its binocular disparity (Born & Bradley, 2005, P. 164). Hippocampal place cells do not respond *only* to place, but also to odors, tactile inputs, recognizable chunks of experiences, and the relative timing of certain events (Itskov et al., 2011;

[21] Though others suggest that wiring length minimization does not strongly correlate with topographic organization (cf Yarrow et al., 2014).

Kraus et al., 2013; Sun et al., 2020; Wood et al., 1999, 2000). Even the neuronal cells constituting the "cortical homunculi", probably the most well known and the most intuitively compelling NSRs, do not always code for single bodily parts (see Penfield and Brodley 1937; Penfield and Rasmussen 1957; Woolsey et al., 1952; Kwan et al., 1978; Wasserman et al. 1992; Schieber, 2001; Aflalo & Graziano, 2006). Indeed, some neurons of the "motor homunculus" appear to code (and control) complex whole-body configurations, in a way that clearly stands in the way of **(1)** (Gordon et al., 2022): if these neurons are vehicle-constituents of the NSRV representing our whole body, they *can't* be representing our whole body without violating **(1)**! All these are significant and systematic (a)-violations of the relevant structural similarity. So, in general, the neat one-to-one mapping from discrete and well-identified "bits" of the neural map to discrete and well-identified bits of the world is a *huge* idealization of the neurobiological reality.[22] As far as neuroscience shows us, (a)-violations are the rule, not the exception, in cortical maps. So it seems that, as a general rule, **(a)** fails to obtain. *A fortiori*, **(1)** does not obtain too.[23]

One could claim that these data pose no threat to **(1)**, as they only show that NSR are much messier than textbook philosophical examples lead us to suppose (thanks to Jonny Lee for this objection). But these data do not "just" complicate the picture. They complicate the picture in a way that directly threatens the obtaining of **(1)** by showing that the relevant vehicle constituents do *not* map onto target constituents in the desired manner. They *don't* show that **(1)** obtains, but in a much messier manner than textbook examples indicate. They show that **(1)** does not obtain.

One could further claim that these data pose no threat to **(1)** because structural similarities between vehicles and targets need not be perfectly accurate nor total. Imperfect, partial, distortive similarities are sufficient to satisfy **(1)** too (cf. Williams & Colling, 2017; Shea, 2018, pp. 140–142). And indeed, sometimes *distortive* similarities might be more functional than non-distortive ones: think the way in which maps of underground metros are way more readable when they do *not* display the actual distance holding between the various metro stations. I think this is an important claim that gets something right. However, I still think that, in the present context, it is insufficient to rescue **(1)**.

For, appealing to approximate similarities allows **(1)** to tolerate *local* (a)-violations and/or (b)-violations, *global* (a)- and/or (b)-violations can't be tolerated. A map can tolerate a (a)-violation (e.g. representing Rome and Paris with a single point) only if it correctly represents other places (e.g. because it represents Lyon and Florence as two distinct points, the former north of the latter). Else, it ceases to be a map in any recognizable sense. And the same goes for (b)-violation. Thus, (a)- and (b)-violations cannot be global. In the case at hand, however, the (a)-violation seems to be if not global at least *extremely* widespread. Neurons responding (and mapping to) single targets, if they exist, are rare exceptions—so rare, indeed, that, to my knowledge, they're yet to be found.

[22] Penfield was explicit on this point. He considered his homunculus as "a cartoon of representation in which scientific accuracy is impossible" intended to be used as an "aid to memory" (both quotes from Penfield and Rasmussen 1950, p.56).

[23] As an aside, notice that the same state of affairs prevents us from considering these neurons and neuronal regions *indicators* in any straightforward and intuitive way.

Perhaps one could argue that, unlike cartographic maps, cortical maps might tolerate *global* (a)- and/or (b)- violations. After all, neural representations have unique properties, and public representation offers only a limited, and mostly analogical, guidance to the understanding of neural representations (thanks again to Jonny Lee for this objection). Whilst this objection, if successful, would rescue **(1)**, I'm not entirely sure that it makes sense; and I think that even if it were sensical, it could not be accepted.

I'm not sure that the objection is sensical because I'm not sure that there is a real difference between something that satisfies **(1)** while allowing for systematic (a)- and/or (b)-violations and something that simply *fails to satisfy* **(1)**. I really don't have the faintest idea of how that difference could be spelled out and articulated—*prima facie*, something allowing for systemic (a)- and/or (b)-violations is simply something that does *not* satisfy **(1)**. If there is a difference between the two, I challenge the objectior to spell it out in a clear manner.

One may further argue that the data I have shown does not really pose any threat to **(1)**. After all, **(1)** requires only a one-to-one correspondence between (at least some of) the vehicle constituents of **V** and (at least some of) the target constituents of **T**. But nothing *obliges* the relevant correspondence to be causal in nature, or to rest on a form of response selectivity. Indeed, such a focus on response factors may suggest that I'm inadvertently shifting towards an indicator view of representation. Consider, in contrast, a paradigmatic structural representation such as a cartographic map. The various points on the map are in no obvious sense *caused* by what they stand for; and their correspondence does not seem to hold in virtue of any response of the map to the environment. Rather, points on the map correspond to various locations partially in virtue of their position in the map, and partially in virtue of certain representational conventions. So, the fact that neurons in cortical maps respond to many environmental stimuli does not *entail* that such neurons map onto many target constituents in the relevant sense of mapping at play in **(1)**.

The objection above makes a valid point: the relevant one-to-one correspondence in **(1)** *does not* need to be causal, or grounded upon response-related factors.[24] More generally, the content of structural representations does not metaphysically depend on causal, correlational or response-dependent factors. This is one important feature that (supposedly) tells apart structural representations from indicators (cf. Cummins, 1996). Yet, as valid as this point is, it fails to provide any compelling reason to reject my line of reasoning. On the one hand, the fact that the one-to-one correspondence in **(1)** *need not* be grounded in causal factors does not exclude that it *can* be so grounded. On the other hand, *I* don't *need* to claim that that correspondence is grounded in causal factors—and, by extension, that the content of structural representations is even partly metaphysically dependent on causal factors. All I need for my arguments to work is that the one-to-one mapping required by **(1)** can be *discovered* appealing to causal factors. And this is not just an assumption that neuroscientists and philosophers *do* make (see references given in this paragraph); it is also an assumption that it is *safe* to make. For, as reminded in Sect. 3, we have good reasons to take neuronal responses to targets to be vehicles representing these targets. Given this safe and reasonable *methodological* procedure to investigate neural representations, the fact

[24] For an exception to this general rule, see (Isaac 2013).

that *bona fide* constituents of NSRV respond to many different targets suggests that they *do not* map one-to-one onto their targets as imposed by **(1)**. So, to repeat: in the case of structural representations, the content metaphysically depends on a structural similarity holding between a vehicle **V** and a target **T**. The similarity requires a one-to-one mapping from the vehicle constituents of **V** to the target constituents of **T** (cf. point **(1)** in Sect. 2 above). When it comes to neural structural representations, the standard *methodology* to discover such a mapping is that of looking at response properties (i.e. causal/correlational factors). And here response properties suggest, *pace* **(1),** that no one-to-one mapping obtains.[25]

Now, concede (for the sake of discussion) that **(1)** obtains. The representationalist would still be swamped by problems. For, **(a)** and **(b)** partially determine the semantic properties of structural representations, and so their global violation yields degenerate semantic properties—which impede cortical maps to be counted as NSRs, for their possession is incompatible with the semantic transparency that characterizes structural representations. Worse still, these degenerate semantic properties make cortical maps unable to play the causal role that characterizes structural representations. So, we can't really *coherently* accept that NSRVs can allow for systematic (a)- and/or (b)-violations. Let me unpack.

Recall (Sect. 2): each vehicle constituent of a structural representation represents the target constituent onto which it maps. Further, relations holding among vehicle constituents represent relations holding among target constituents: $\Re(V_A,V_B)$ represents that $\Gamma(T_A,T_B)$. But actual neural responses "map onto" more than one target constituent—neurons do not respond *only* to their preferred stimuli. So, in the case at hand, V_A does not map only onto T_A, it maps also onto $T_@$. But then, what does $\Re(V_A,V_B)$ represent? $\Gamma(T_A,T_B)$, $\Gamma(T_@,T_B)$ or some mixture of the two? Defenders of NSRs are unable to answer this question in a satisfactory way: all answers fail to deliver contents with the requested semantic transparency. Moreover, since the content lacks the desired semantic transparency, it is unclear when **V** represents**T**. Hence it is unclear whether **V** is able to play the action guiding role imposed by **(2)**. Since **(4)** is entailed by **(2)**, **(4)** is in danger too.

To see why this is the case, suppose, first, that $\Re(V_A,V_B)$ represents only $\Gamma(T_A,T_B)$—or $\Gamma(T_@,T_B)$. Notice that this is a fairly substantial supposition: it amounts to supposing that **V** actually satisfies **(1)** and so has the required semantic transparency. But even with a substantial supposition in place, it is not yet determined whether it is represented $\Gamma(T_A,T_B)$ or $\Gamma(T_@,T_B)$. The supposition is that only one of the two is represented—but now it is necessary to determine *which one* is represented. Yet, what **V** represents is determined by the relevant structural similarity **V** bears to some target—which *does not* discriminate between $\Gamma(T_A,T_B)$ and $\Gamma(T_@,T_B)$. So, **V**'s content

[25] Notice an objector *cannot* deny this latter methodological point without *thereby* granting my point that NSRV have not been observed. For, in the case of neuronal maps (and other *bona fide* NSRV), it is standardly claimed that the relevant mapping has been discovered through such means. But, if these means were inadequate to observe NSRVs, then it clearly follows *that we've not observed them*—and this is exactly my point!

is indeterminate: Sure, **V** represents one and only one target **T**, but *which individual target T is represented is left entirely open*.[26] **V** is thus semantically transparent in name only. Further, since **T** is indeterminate, whether **(2)** and **(4)** obtain is left entirely unclear. If we don't know what **V** is structurally similar to, we can't determine whether increasing (or decreasing) that similarity increases (or decreases) the agent's chance of success. **V** would thus be a vehicle of a structural representation in name only. Moreover: the fact that neurons commit systematic (a)-violations is functionally relevant—it improves the way in which our neurocognitive mechanisms work (Chakrabarty & Martin, 2000; Martin et al., 2005). If the way in which such mechanisms function really is best explained representationally, a representational explanation should should be expected to *emphasize* that fact, rather than hiding it under the carpet assigning these representational vehicles a single representational content *by fiat*.

So, in the case at hand, a representational explanation should *not* choose one between $\Gamma(T_A,T_B)$ and $\Gamma(T_@,T_B)$—it should find a way to say that both are represented. Suppose, then, that $\Re(V_A,V_B)$ represents *both* $\Gamma(T_A,T_B)$ and $\Gamma(T_@,T_B)$. So, $\Re(V_A,V_B)$ has a composite content, which might be expressed by $\{\Gamma(T_A,T_B)$ & $\Gamma(T_@,T_B)\}$. But clearly such a content is not semantically transparent in the desired way. But the desired semantic transparency seems entailed by **(1)**, and so now it seems that **(1)** is not the case. This conclusion generates a contradiction—in fact, we're trying to determine what would **V**'s content be, supposing that **(1)** obtains in spite of the various (a)-violations it suffers from. And even leaving this problem aside, there would be problems with **(2)** and **(4)**. Suppose that an agent is using the representation **V** (including $\Re(V_A,V_B)$) to guide their behavior in respect to a **T** such that $\Gamma(T_A,T_B)$ is the case but $\Gamma(T_@,T_B)$ is not the case. Here, it is legitimate to expect the agent to non accidentally *succeed*: $\Re(V_A,V_B)$ carries information about $\Gamma(T_A,T_B)$ which the agent can "use" to appropriately orchestrate their behavior. But if $\Re(V_A,V_B)$ actually represents $\{\Gamma(T_A,T_B)$ & $\Gamma(T_@,T_B)\}$, then it is false (or extremely non-accurate). The truth value (or degree of accuracy) of **V** no longer correlates with the agent's behavioral success, and so **(2)** fails to obtain. Given that **(4)** is entailed by **(2)**, **(4)** fails to obtain too. Of course, one could solve *this* specific problem by arguing that the composite content is something that could be best expressed by $\{\Gamma(T_A,T_B)$ or $\Gamma(T_@,T_B)\}$. But now the content of **V** is plainly disjunctive, and falls prey to the disjunction problem in its various forms (cf. Neander, 2017). And notice that, since the original assumption was that systematic (a)- (and (b)-)violations are admissible, the disjunction here seems unrestrained.

[26] A tempting and obvious solution to this problem is that of resorting to a form of informational (or information-based) semantics; that is, claiming that each neuron "maps onto" the stimulus about which it carries the most information (cf. Wiese 2017, pp. 219-223, also (arguably) Piccinini 2020b). However, such informational linkages seem unable to ascribe determined contents (Artiga & Sebastian 2018; Rosche & Sober 2019). More generally, theories of structural representations interact poorly with informational accounts of content (cf. Facchin 2021a). A second solution is that of appealing to the agent's actual context (Ramsey 2007). But this solution can only work in *some* cases of successful online behavior. If the relevant vehicle is used in a decoupled manner, in service of offline cognition, then there is *nothing* in the agent's context that can discriminate between $\Gamma(T_A,T_B)$ and $\Gamma(T_@,T_B)$—else, the agent' would not be decoupled from at least one of them. So, the solution does not generalize and fails to appropriately restore content determinacy. Other solutions are far less obvious, and thus cannot be considered here.

Taking stock: that the neurons of neural maps do not map in a neat one-to-one fashion onto stimuli is a serious problem for the defender of NSRs. The absence of the required one-to-one mapping may be enough to claim that neural maps fail to satisfy **(1)**. And, even accepting that the absence of such a map is no reason to deny that **(1)** obtains, there would still be significant problems with **(2)** and **(4)**. It would be at best unclear whether neuronal maps guide their "users" actions in the way structural representations are supposed to carry out their action guiding duties.

Defenders of NSRs might then be tempted to abandon **(2)** and **(4)** to secure the status of cortical maps as NSRVs. But this is unwise. Recall why NSRs are central in the "cognitive neuroscience revolution". They are central because they allow for the happy marriage of mechanical and representational explanations (Sect. 1). NSRs allow for this marriage because their NSRVs—the causally efficacious bits and pieces that operate within our neurocognitive mechanisms—are imbued with content: their physical shape has important semantic properties in a way such that these semantic properties are allowed to play an active causal role within our neurocognitive systems (Sect. 2). In the case of NSRVs, then, the semantics itself does the pushing and pulling required by mechanistic explanations. But, assuming that representational accuracy is conducive to pragmatic success, this view *entails* **(2)**: the degree of accuracy between vehicle and target must be reflected in the agent's odds of pragmatic success. So, abandoning **(2)** means either (i) abandoning the view that representational accuracy is conducive to pragmatic success or (ii) abandoning the view that the content of NSRs plays a causal role compatible with it being a part of mechanistic explanations. Both options are unattractive to the defender of NSRs. Denying (i) is tantamount to admitting that representations are conducive to success regardless of their truth or accuracy value—which is clearly false. But denying (ii) amounts to conceding that the relevant semantic properties of NSRVs do not play any mechanistically relevant causal role—de facto undermining the theoretical attractiveness of NSRs for cognitive neuroscience in general and *mechanistic* cognitive neuroscience in particular (cf. O'Brien, 2015; Williams & Colling, 2017).

In spite of their appearance, then, neural maps are not NSRVs—the structural similarity that they so obviously boast (to our eyes) might not even be really present. And, even if it were present, it would not play the required representational role.

3.3 Activation spaces

Thus far, I've in an important sense considered only *single* responses, either of individual neurons (Sect. 3.1) or of multiple neurons topographically organized in neural maps (Sect. 3.2). Some defenders of NSRs would claim my focus has been too narrow. To see NSRVs one should look at *multiple* responses from a single neuronal structure. For, the relevant (i.e. NSR-underpinning) similarity does not hold between a single activation and a target. Rather, it holds among the structure's entire activation space (i.e. set of all possible responses) and the entire target domain (i.e. the set of all targets the structure is sensitive to). As far as I can see, there are two different arguments for this claim.

The first—and more widespread—variant is ultimately based on the analysis of the behavior of a large class of neurocomputational models (cf. Churchland, 1995; O'Brien and Opie 2004; Grush, 2004; Shagrir, 2012; Williams, 2017; Wiese, 2016, 2017).[27] Shagrir (2018) usefully expresses the idea common to all these arguments in terms of *input–output modeling*. Let f be the function relating the inputs and outputs of a neurocomputational model. In Shagrir's view, such a model is a model of a target domain **T** if, when V_A and V_B are in the relevant input–output relation specified by f, then the corresponding elements in the target domain (T_A and T_B) stand in a relation mathematically described by f too. Consider, for example, a model **M** that takes as input velocities and yields as outputs space traveled in a minute at that velocity. According to Shagrir, **M** is an input–output model of its target domain **T** just in case it multiplies the input value by 60—given that $s = vt$ and here $t = 60$ s. When this happens, the activation space of **M**—that is, the set of all **M**'s input–output pairings—is clearly structurally similar to the target domain, in a way that seemingly vindicates **(1)**. What, then, about **(2)**-**(4)**? The argument to the effect they obtain vary from account to account—but here I will ignore them, as they won't play any role in my argument below.

The second—and less widespread (to my knowledge, is made only by Williams & Colling, 2017)—argument is based on a technique to analyze neuroimaging data known as *representational similarity analysis* (RSA, see Kriegeskorte et al., 2008). RSA belongs to the family of "neural decoding"—or, more soberly, multivariate patterns analysis—techniques. These techniques operate on various types of imaging data to investigate neural representations (e.g. Haxby et al., 2001).[28] RSA typically operates on voxels—think of them as three dimensional pixels "making up" the images—and their activation levels. Each activation is treated as a vector of voxels activation levels, so as to compute the distance (i.e. dissimilarity) between each pair of vectors. Based on these distances, the activations are arranged in a *representational dissimilarity matrix*: an activation space expressing the dissimilarity between each pair of activation as a scalar quantity (i.e. a single number; see Kriegeskorte & Kievit, 2013 for an accessible introduction to RSA). Importantly, the pattern of similarities and dissimilarities between neuronal responses revealed by the representational dissimilarity matrix "mirrors" the pattern of similarities and dissimilarities expressed by subjects in their similarity judgments (cf. Connolly et al., 2012; Ritchie et al. 2014; Carlson et al., 2014). So, if two responses are similar (i.e. $\aleph(V_A,V_B)$), then their two targets are similar (i.e. $\Gamma(T_A,T_B)$), in a way that seemingly vindicates **(1)**.

Sadly for the defender of NSRs, however, none of these two arguments establishes that **(1)** obtains. Although *both* arguments show a structural similarity holding, they show it holding amongst the *wrong* sorts of things—at least, given the characterization of structural representations they endorse (Sect. 2).[29]

Condition **(1)** requires a structural similarity to hold between *a representational vehicle* **V** *and a represented target T*. But the structural similarities shown above do

[27] See also (Rutar et al., 2022) for a more nuanced—and less structural-representationalist—treatment.

[28] Pitched at this level of generality, the claim is importantly contested (cf. Ritchie et al., 2019; Gessel et al.2021). These critical arguments, however, do not apply to RSA, and so I will ignore them here.

[29] I will make a more general point about this issue in the *post scriptum* of this paper.

not hold amongst *individual representational vehicles* and *individual targets*. This is especially obvious in the case of the first argument based on input–output modeling. In that case, the structural similarity holds between a *computational process* pairing inputs and outputs and a certain environmental process. But whilst environmental processes can be represented targets, computational processes can't be representational vehicles. Indeed, on a number of standard accounts of physical computation, computational processes are perspicuously (i.e. informatively, non-circularly) *defined over* representational vehicles (cf Fodor 1981; O'Brien & Opie, 2009; Maley, 2021a).[30] This clearly implies that computational processes and representational vehicles are distinct: the representational vehicles are the primitives used to define computational processes, and computational processes consist in the (rule-based) manipulation of vehicles. Compare: shuffling is a process we can define over certain primitives (for example, the cards of a deck) whereby such cards are manipulated (chiefly, by modifying their placement within the deck in a quasi-random manner). Just like the process of shuffling is distinct from the cards it is defined over, so too computational processes are distinct from the representational vehicles they are defined over. Thus, even if input–output modeling were observed in the brain (something that, to my knowledge, defenders of the "cognitive neuroscience revolution" haven't yet argued for), that would not be equivalent to observing any NSRV. For, although in this case a structural similarity *would* be observed, that structural similarity *would not* hold among a single representational vehicle and its target, and so it would not satisfy **(1)**.

RSA suffers from a similar problem, though in an attenuated (and less obvious) form. As pointed out by (Davis & Poldrack, 2013; see also Coraci, 2022 for a philosopher-friendly analysis) it is not entirely clear whether the structural similarity RSA reveals depends on the *representations* involved within a cognitive process or on the *cognitive process* being run during the experimental trial. Consider, for instance, the neuronal representation of a male face smiling and the neuronal representation of a female face smiling. These two neuronal activations are likely similar because they represent similar things. But now consider the neuronal activation involved in representing a smiling face and a puppy. These neuronal activation might be similar—but, if so, their similarity would not be due to the similarity of their *contents*, but rather to the fact of a same *cognitive process* (say, judging both the smiling face and the puppy good and having a positive affective response to them) operates on them both. These two scenarios can be disentangled with certain appropriate experimental procedures. But the need to disentangle them weakens any inference from the structural similarities shown by RSA and the claim that **(1)** obtains.

Worse still, even when the structural similarity revealed through RSA techniques is due to the similarity of the representations (rather than the processes), that structural similarity still fails to support the claim that **(1)** obtains. For, the relevant similarity holds between a *representational dissimilarity matrix* and various targets. But representation dissimilarity matrices are not neural vehicles: not only do they abstract away from the spatiotemporal information that is needed to identify vehicles (see Haxby et al., 2014, p. 439; Kriegeskorte & Diedrichsen, 2019, p. 418). They're not realized

[30] Or non-representational computational states more generally (cf. Piccinini 2015).

by neurons and their activities, nor are they tokenized in our heads. They're not what vindicating **(1)** requires in this context.

Defenders of NSRs could plausibly object that, whilst I'm correctly pointing out that computational processes and representational dissimilarity matrices are not neural vehicles, they still *reveal something important* about our neural vehicles, and can be plausibly and fruitfully used as proxies of the latter (Roskies, 2021).

This objection gets an important point right: the structure of computational processes and analysis techniques such as RSA *do* reveal important pieces of information about our neural activations and neural vehicles (if they exist). But still, to correctly interpret the information they reveal, we should first grasp clearly what they are and how they operate. And doing that, I submit, prevents us from claiming that any structural similarity they reveal satisfies **(1)**. For example, whilst RSA reveals that activation spaces are structurally similar to the stimulus space, it is only a *model* of neural activity. And in this model, the *bona fide* neural vehicles are the individual points that populate the space, rather than the space itself:

> "The dominant theoretical underpinning of representational analyses in most content areas of *f*MRI research is that stimulus *representations can be thought of as points in an n-dimensional space*. This characterization of neural representations in terms of n-dimensional spaces follows from influential work in cognitive psychology on how *psychological representations can often be characterized as points in a representational space* and how a variety of cognitive processes, such as stimulus generalization, categorization, and memory, can be modeled as geometric operations on these representations." (Davies & Poldrack 2013, p. 109, emphasis added).

Thus, a proper understanding of RSA *prevents* us from using any structural similarity it eventually reveals to substantiate claims to the effect that **(1)** obtains. An isomorphic point could be made in regards to computational processes. Computational processes are (standardly) *defined over* representations. And so, any structural similarity between computational processes and targets is by definition unsuitable to vindicate **(1)**.

But can't defenders of the "cognitive neuroscience revolution" somehow claim that entire activation spaces (or entire computational processes) are representational vehicles, so at allow **(1)** to obtain?

No, they cannot. The reason is simple. Activation spaces and representational dissimilarity matrices show us that—for example—if two neuronal responses are similar, then their targets (i.e. what these responses are responses to) are similar too. Rewriting this in the notation used throughout the paper, the result is: $\mathfrak{R}(V_A,V_B) \rightarrow \mathfrak{R}(T_A,T_B)$.[31] The same notation applies to computational processes: if V_A and V_B stand in a mathematical relation described by a function f, so too do T_A and T_B. Again, $\mathfrak{R}(V_A,V_B) \rightarrow \mathfrak{R}(T_A,T_B)$. Note that, in both cases, individual neuronal responses (and individual computational states) are treated as the material constituents of some larger vehicle—the entire activation space, or the entire computational process. The problem,

[31] Notice that I'm writing "$\mathfrak{R}(T_A,T_B)$" for the relation upon which the structural similarity is based is the same on both sides of the mapping.

however, is that they *can't* be material constituents of a larger vehicle, especially not in a mechanistic framework. As Kirchhoff (2014, 2015) has aptly noticed, material constitution is typically taken to be a *synchronic* relation holding between the constituents and the constituted entity. The mechanistic conception of constitution agrees, and actually relies on it to tell apart constitutive from merely causal relations within mechanisms (see, e.g. Krickel, 2018; Baumgartner et al., 2020).[32] So, if vehicle constituents $V_A \ldots V_N$ constitute vehicle **V** at time *t*, then $V_A \ldots V_N$ must all be present at *t*. Yet, in the case at hand, we haven't observed the vehicle constituents being all present at the same time. And indeed, they *can't* be present at the same time. In implemented computational systems, inputs *precede* in time their outputs for a quite obvious reason: the presentation of the input must cause (together with the relevant computational internal states) the tokening of the output at a later time step.[33] And when it comes to activation spaces, it is important to notice that they chart various different activations of a *single* neural region of interest. But a region cannot tokenize *different* activations at the same time. It can only tokenize them *in sequence*, at different times. As a consequence, neither the individual neuronal responses "making up" an activation space nor individual computational states can be rightfully considered as material constituents of a larger vehicle.[34] Notice also how this observation ties a loose end hanging from (Sect. 3.1). At the end of that section, I argued that, even if the increasingly popular view that indicators are structural representations were correct, that wouldn't entail that individual responses of individual neurons are representational vehicles structurally similar to their targets. At best, they are vehicle constituents—or so I conceded back then. It is now possible for me to retract that concession. Different individual neuronal responses of individual neurons *can't* be tokenized at the same time, and so they *can't* be material constituents of a larger neural vehicle. Thus observing them in *no* way amounts to observing a NSRV.

At this juncture, a defender of NSRs may claim that my arguments overlook the fact that many allow for structural representations to be "made up" by more than a representational vehicle. For example, Shea (2018, p. 118) defines structural representations as: "A collection of representations in which a relation on representational vehicles represents a relation on the entities they represent". So do other defenders of structural representations, including Swoyer (1991), Ramsey (2007) and, arguably, Cummins (1989). So, **(1)** need not be narrowly defined in terms of *single* vehicles, as I did in (Sect. 2). And if so, then the structural similarity shown by activation spaces and representational dissimilarity matrices can satisfy **(1)**.

[32] In all fairness, some philosophers try to elaborate a *diachronic* account of constitution (see. Leuridan & Lodewyckx 2021; Kirchhoff and Kiverstein 2021; Kiverstein and Kirchhoff 2023) which may be used to counter my point. I'm skeptical about these accounts, and I would wedge against them a modified version of Krickel's (2023) objection. But I can't articulate it here. So, I will only notice that defenders of the "cognitive neuroscience revolution" do not seem to be interested in such accounts, in a way that makes their view vulnerable to my objection.

[33] Of course, the same may not be true of non-implemented (purely mathematical) computational systems. But looking at such abstract entities could hardly allow us to observe *neural* representational vehicles.

[34] As a reviewer noticed, this also prevents defenders of the "cognitive neuroscience revolution" from categorizing inner simulations as structural representations, as they arguably should. A problem more for the cognitive neuroscience revolution.

This move, however, is unwise. As (Cummins, 1996) noticed, in the case of structural representations, the individual material constituents are *not* representations in their own right. Their representational status of the parts depends on the status of the whole. A vehicle constituent V_A *does* represent T_A—but only because it is a part of a larger vehicle **V** representing **T**. The representational powers of V_A thus piggyback on the representational powers of the entire **V**. A definition (or characterization) of structural representations in terms of multiple representations *obscures* this fact, and should thus be resisted. Worse still, the move would generate problems with mechanistic explanations. For, suppose that structural representation can be made up by "collections" of individual representational vehicles. What would these complex representations be, in the case at hand? They would be *abstracta*: sets of neuronal responses (or computational states) that are structurally similar to sets of worldly stimuli (or states). But sets—and abstracta more generally—c*annot* be components of mechanisms. Mechanisms and their components are always concrete (cf. Craver, 2007). So, redefining structural representations in terms of multiple vehicles does not actually help the cause of the "cognitive neuroscience revolution".

Am I suggesting that the structural similarity displayed by activation spaces and (some) computational processes is necessarily representationally idle? Not necessarily. I'm only denying it holds between vehicles and targets so as to underpin NSRs. But it can still have some relevant representational role. For example, it might determine the content of some other type of representation. There are various theories of content based on structural similarity (e.g. Cummins, 1996, O'Brien and Opie 2004)—and while these theories often focus on the structural similarity between individual vehicles and targets, nothing prevents us from applying the same idea to *multiple vehicles* and *target domains*.[35] On this view, individual vehicles would get their content in virtue of the structural similarity holding between a set of different vehicles and a target domain. Each vehicle would thus represent what it represents in virtue of its overall role in the similarity. This intuition could be refined in a full-blown theory of content—but doing so is a task for another paper to carry out. But notice that, even if such a theory of content were provided, it *would not* lend support to the claim that activation spaces/neural dissimilarity matrices/multiple neuronal responses are structural representations. There is a clear and obvious difference between a set of vehicles being structurally similar to a set of targets and individual vehicles being structurally similar to individual targets. The former *just isn't what (1) requires*.

3.4 Alternative neural vehicles

Whilst neuronal responses are the *main* neuronal vehicles cognitive neuroscience is interested in, they're not the *only* vehicles cognitive neuroscience is interested in. So, what about *those*? Do they underpin NSRs? No they don't, and for fairly obvious reasons.

[35] Indeed, Churchland's (1992) original structural similarity-based account of content was explicitly focused on *multiple* vehicles.

Neuronal connections have often been considered representational vehicles. Indeed, connectionists have long argued that connections between neurons may encode information, functioning as our long-term semantic memory (cf. McClelland et al., 1986). However, it is commonly accepted that if connections encode information, they do so in a highly distributed way: single connections store multiple "bits" of different contents, and single contents are "spread over" many connections (see Van Gelder, 1991; Grush & Mandik, 2002). But if this is the case, if really multiple contents are *simultaneously* encoded by *many overlapping connection*, then clearly the mapping from vehicle constituents to target constituents is *many-to-many*; and so **(a)**—and, *a fortiori* **(1)**—fail to obtain for reasons connected with systematic (a)-violations explored in Sect. 3.2 (see also Facchin, 2021a for a different argument to the same effect). So, if connections are representational vehicles (which is disputable, see Ramsey, 2007), then they're not NSRVs.

Some neuroscientists have recently suggested that global brain states are neural vehicles that represent the agent's overall state (Kaplan & Zimmer, 2020; Westlin et al., 2023). As far as I can see no one has ever claimed that global brain states are NSRVs. And it is indeed hard to see how they could underpin NSRs: there's clearly no decoupling from an agent's current state! So, global brain states clearly fail to satisfy **(3).**

Lastly, Chemero (2009) and Martinez and Artiga (2021) have argued that *neuronal oscillations* (i.e. patterns of time-locked neuronal activity, see Buzsaki, 2006) are representational vehicles. Are they NSRVs? To my knowledge, no one has yet articulated this view. So, I can't provide a detailed analysis of it. However, there are potent *prima facie* reasons to provide a negative answer. Firing patterns instantiated in *different* times can't be constituents of a single vehicle (see Sect. 3.3), and this seems to prevent many neuronal oscillations from qualifying as NSRVs. Further, the individual neuronal responses "making up" the oscillations would still fail to map on individual targets as seen in (Sect. 3.2), generating all the problems discussed in that section.

Are there other potential neural vehicles? Not to my knowledge. Sometimes neuroscientists talk about entire neural structures representing (e.g. the fusiform face *area* is sometimes said to represent faces) but it seems clear that it is a metonymic way of speaking: what neuroscientists most plausibly actually mean is that the *responses* or *activations* in various structures represent things. And, there seems to be no other candidate vehicles. Of course, I cannot exclude that new, more sensitive experimental techniques will reveal functionally salient neuronal aggregations *between* the level of the single neurons and that of neuronal maps, or below the level of individual voxels. These may qualify as NSRVs. But surely such vehicles have yet to be identified—so, we can neither observe them right now, nor can they provide a reasonable ground for the "cognitive neuroscience revolution" *right now*.

3.5 Neural representations unobserved

Time to take stocks, and summarize this long section. I have argued that NSRVs have not been observed or manipulated.

In (Sect. 3.1) I focused on individual neuronal responses. I argued that the claim that individual neuronal responses are NSRVs is ambiguous, as it admits three different readings. No such reading, however, allows neuronal responses to break down into interrelated constituents in the desired manner; and indeed the claim that a constituent of an individual neuronal response represents a constituent of the response's target would be a *reductio* of the idea that individual neuronal responses count as NSRVs. Individual neuronal responses, I suggested, are more plausibly interpreted as indicator representations.

But aren't indicators a special case of structural representation, as some philosophers argue? I am officially neutral on this issue (at least in this paper). Yet I have noticed that, even if they were, that would not "turn" individual neuronal responses into NSRVs. So, even in this case, NSRVs remain unobserved—at least at the level of individual neuronal responses.

In (Sect. 3.2) I focused on neuronal maps, claiming that they are not NSRVs. First, I have argued that the topological structural similarity holding between neuronal maps and their target domain does not satisfy **(2)**. Contrary to what **(2)** requires (a)- and (b)-violations of that structural similarity do not decrease an agent's odd of non-accidental success. I also considered other possible structural similarities tying together neuronal maps and their targets, which, not being based on their apparent topological similarity, would be impervious to the argument above. I then ruled this possibility out based on the fact that individual neurons (that is, the relevant vehicle constituents) *do not* map one-to-one onto their targets as required by **(a)**, and so, **(1)** systematically fails to obtain.

In (Sect. 3.3) I focus on activation spaces. I claimed that, whilst such spaces actually are structurally similar to their target domains, that structural similarity does not satisfy **(1)**. In fact, **(1)** requires a structural similarity holding between individual vehicles and targets—but activation spaces just cannot be coherently considered to be individual vehicles.

Lastly, in (Sect. 3.4) I considered a number of other alternative neural vehicles that might underpin NSRs, showing that none actually underpins them for fairly obvious reasons.

A final word of clarification. Above, I have considered a relatively small number of case studies. One might thus worry that my arguments are fueled by little data to *inductively* support my conclusion that NSRVs have not been observed. Fair point, but my arguments here are not *inductive* arguments. I'm not claiming that individual neuronal responses, neural maps, and activation spaces all likely fail to satisfy **(1)**-**(4)** because many of them fail to satisfy them. Rather, my arguments show that these entities *cannot* satisfy **(1)**-**(4)** for various *principled reasons*, and thus that they can't *in principle* qualify as NSRVs.

4 Objections and replies

Supporters of the "cognitive neuroscience revolution" will no doubt wish to resist my conclusion. Here I consider some intuitive objections to resist it, showing that they do not really work.

Objection #1: The arguments in (Sect. 3) have focused on various types of neuronal responses. But when it comes to discussing NSRs, that focus is misguided. For, the structural notion of representation has been crafted *as an alternative* to the indicator, response-based one (cf. Cummins, 1996; Ramsey, 2007; Williams & Colling, 2017). According to the structuralist view of representation, neural structures function as representations not because they *selectively respond* to something else, but because they model, or provide a "map" of, something else. According to this objection, then, I have systematically looked at the *wrong* sort of neural structures. No wonder I have failed to observe NSRVs!

Response: There is something to the objection. Even if some philosophers dispute that there is an actual difference (see Sect. 3.1) structural representations and indicators are typically presented as alternatives. It is also true that, unlike indicators, structural representations are supposed to function as representations not *because* of how they respond, but because of how they model or "map" some target. That being said, as already hinted at in (Sect. 3.2), *nowhere* in this paper I have claimed that the structures examined in Sect. 3 are representations because of how they *respond* to their targets. Sure, I have discussed various cases in which the structural similarity between a (candidate) NSRV and its target **T** has been *discovered* thanks to the response properties of the candidate vehicle. But that *modus operandi* is entirely compatible with the claim that the candidate NSRV acquires its representational role not because of how it *responds* to **T**, but because of how it *models* **T**. In other terms, one can accept that looking at the response properties of a candidate vehicle **V** can be used to "discover" the structural similarity tying it to **T** without thereby having to commit to the claim that **V** represents **T** *because of how it responds* to **T**. And indeed, condition **(3)** of the definition offreded in Sect. 2 avoids precisely such a commitment: since structural representations must be able to function offline, the response profile can't be what *makes* something into a structural representation. This, however, clearly does not exclude that structural representation can be used *online*, and tokenized by means of neuronal responses!

As a further point in response to this first objection, notice that, even if the objection were not misguided, it would not really hinder the claim I'm defending here. For, as noticed at the beginning of Sect. 3, looking at neuronal responses is the standard way to try and observe neural representations (structural or otherwise). Even defenders of the "cognitive neuroscience revolution" adopts this procedure to claim that NSRs have been observed (cf. Piccinini 202a; Thomson & Piccinini, 2018). So, if this procedure were unable to make us observe NSRs, it follows that, well, we *can't* have observed them yet—and so that *we haven't observed NSRs yet*. But that is exactly my claim! So, even if objection #1 were on the right track, we would not be left with a refutation of the claim I've been defending here. We would be left with the need of finding new, more promising ways to try and observe neural representations. A need, I submit, whose satisfaction falls squarely on the objector's shoulder.

Objection #2: The account of structural representations in Sect. 2 is too *demanding*. A less demanding account would reveal that NSRVs are not just present in our brains, but that they have indeed been observed.

Response: Two points in reply. First, we lack an alternative, less demanding, account of NSRs. The account in Sect. 2 is widely used (see, for example, Wiese, 2016, 2017;

Williams, 2017; Lee, 2019), and the (few) alternative ones are not less demanding—indeed, they're often *more* demanding, as they adopt a *stronger* reading of **(1)** in terms of homomorphisms. Lacking any less demanding alternative, the objection is pretty toothless.

Secondly, it's hard to even imagine the shape of a less demanding alternative. Presumably, the alternative should discard or weaken at least one condition among **(1)**—**(4)**. But my reading of **(1)** is already the weakest one acceptable, and **(1)** cannot be discarded without thereby discarding the very idea of a *structural* representation. My reading of **(3)** is also the weakest reading of decouplability on offer (cf. Chemero, 2009, pp. 55–65; Gładziejewski, 2015); and **(3)** cannot be discarded either, as decouplability is an *essential* feature of representations (Haugeland 1991; Orlandi, 2020). **(2)** *could* be weakened and discarded—but doing so would hinder the causal relevance of content, in a way that hinders its relevance in mechanistic explanations. Defenders of NSRs can't thus rely on this move—at least, not without abandoning their mechanistic commitments. And since **(2)** entails **(4)**, **(4)** seems off limits too.

Objection #3: NSRs are *action-oriented* representations (Piccinini, 2022; Williams, 2017).[36] So, they don't represent the world objectively, but in action-salient terms. But what can "representing the world in action salient terms" mean, if not that the world is represented in a somewhat *distortive* way, which emphasizes action-relevant features at the expense of action-irrelevant ones? But if the world is represented in such a distortive way, the representation must be somewhat false or inaccurate—in a way that is *nevertheless conductive to an agent's behavioral success* (cf. Tschantz et al., 2020 for a proof of concept). But this clearly runs counter to **(2)**—for **(2)** establishes that there's a direct proportionality between the accuracy (or truthfulness) of a representation and the agent's odds of behavioral success. As a consequence, **(2)** should be discarded—and with it, all the arguments above that hinged on **(2)** failing to obtain (cf. Section 3.2). So, NSRVs have been observed, after all.

Response: The objection misconstrues the sense in which action-oriented representations are distortive. Sure, they do not represent the world "as is" (whatever this means)—but that's not to say that they represent it falsely or inaccurately. They represent it *through a pragmatic lens*, and what is represented through such a lens can be either accurate/true or inaccurate/false. If I represent a 6 kg stone as throwable, I'm accurately representing the stone in an action oriented manner. If I represent a 666 kg stone as throwable, I'm inaccurately representing it in an action-oriented manner. Compare: if, by looking through red glasses, I see clouds being red, I'm *not misperceiving*—I'm accurately perceiving through red glasses. Thus, the action-oriented nature of NSRs does not force a rejection of **(2)**—or of the "bits" of my arguments based on **(2)** failing to obtain.

Objection #4: The argument in (Sect. 3.2) is a bit too quick in establishing that individual neurons map onto *many* targets in a way that poses a problem for **(1)**. Neurons need not represent each target to which they respond. Taking a page out of Dretske's (1988) book, one could argue that individual neurons have the *function to represent* only one target, plausibly their preferred one. That might be enough (or at

[36] On the concept of action oriented representations, see (Clark 1997). Curiously, Clark's original example of an action oriented representation is that of Mataric (1991) "spatial map"—a robotic replica of the "spatial map" in the rat's hippocampus. So, it seems that action oriented representations were NSRs all along.

least a substantial step towards) solving the problem with **(1)**, in a way that also avoids the problems with **(2)** neurons mapping onto many targets generated.

Response: Whilst taking a page out of Dretske's book would solve these problems, the defender of NSRs *can't* rely on Dretske's solution. Dretske assigns functions only after a *learning period*, which stabilizes the function (i.e. determines what the neuron is "supposed to" represent). But real brains have no learning period separate from a non-learning period. Neurocognitive networks are constantly re-organizing and can quickly learn to operate in very odd conditions. As enactivists have repeatedly pointed out, our sensorimotor system can *learn* to operate even in conditions under which sensory and motor signals have been dramatically altered—for exemple, due to one's usage of "inverting goggles" (Hurley, 1998; O'Regan, 2011). Surely a neuron's learning phase should be over well before the subject is old enough to take part in psychological experiments involving the usage of "inverting goggles"! More generally, it is extremely tricky to assign *well-defined, individual* functions to neuronal areas. Neural functions appear to be multiple, multidimensional, not well-determinate and extremely context dependent (cf. Anderson, 2014; Burnston, 2016; de Wit & Matheson, 2022)—and so will be the contents they ground. Yet, as seen in (Sect. 3.2), NSRs require reasonably well-determinate contents to function.

Objection #5: My objections to NSRs were hyper-focused on the features of their vehicles. Yet structural representations need not reside at such an "implementational" level of abstraction. They may reside at a higher, "algorithmic" level, and dangle free of implementational details[37] Cummins (1989), for example, situated them at the level of program execution. Johnson-Laird (1983) thought of his mental *models* as existing roughly at the same "algorithmic" level of abstraction. Similarly, Danks's (2014) suggestion that cognitive representations are graphical *models* sits at a level of abstraction more akin to that of program execution than the implementation level. My arguments are silent about these structural representations and their neural vehicles. So, it fails to rule out NSRs at higher levels of abstraction.

Reply: My reply is simple. Given the relevant characterization of structural representations defenders of cognitive neuroscience revolution accept (Sect. 2), structural representations are characterized in terms of their *vehicle* properties; that is, in terms of the *concrete material thing doing the representing*. And such a concrete, material thing presumably sits at the "implementation" level. So, if one's structural representations "dangle free" from any implementational details—or leave them unspecified—then *they are not structural representations in the relevant sense*, and their existence, even if proved, *would not* support the cognitive neuroscience revolution. This shouldn't be surprising. Mechanistic account of neuroscience have always argued that the "algorithmic" and "computational" level of explanations are not autonomous from the nitty–gritty implementational level, and that these levels of explanation provide at best *sketches* of mechanisms in which the implementational details still have to be filled in (Piccinini & Craver, 2011; Van Bree, 2023).

[37] But see Maley (2021b) for an argument to the effect that, in the case of analog representations (including structural ones) the difference between implementational and algorithmic level collapses.

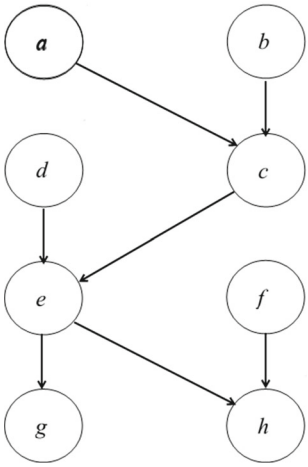

Fig. 3 A graphical model capturing the statistical dependency relation of some random variables. Drawing by the author

Of course, this *does not* mean that structural representations must be characterized at the finest possible grain of neurological detail (e.g. in terms of the release of single neurotransmitter molecules). They may be identified at higher levels of abstraction—say, for example, at the level of individual neuronal responses (Sect. 3.1), cortical maps (Sect. 3.2) or at the level of multiple responses (Sect. 3.3). And they may even be identified on the basis of an algorithmic (or even computational) description of such levels of abstraction. But to reveal genuine structural representations—at least given the characterization in (Sect. 2)—these higher levels of abstraction *must* somehow be mapped onto the physical vehicle doing the representing. Else, we risk wrongfully considering non-structural types of representations as structural representations only because we *model these representations* using an iconic format. Consider, for example, the graphical model in Fig. 3

Figure 3 represents a simple "Bayesian model" (i.e. a directed acyclic graph), which can be used to model a target phenomenon **T**. Now, the model—as it is presented to us—surely *seems* a structural representation of **T**: the nodes *a*-*h* map one-to-one on aspects of **T**, and the pattern of arrows "recapitulates" the statistical dependencies in **T**. But notice that the arrows and nodes we see are not the vehicle underpinning the model—the vehicle is a complex series of voltages (at the level of the implementation) or "0"s and "1"s (as a higher level) somewhere in my computer.[38] And there is no guarantee (nor any reason to believe) that these voltages (or "0"s and "1"s) will be structurally similar to **T**. Further, the impression of iconicity can be easily dispelled by visualizing the model of Fig. 3 in a less graphical (pun intended) format—for example, as the probability distribution p(a, b, c, d, e, f, g, h) = p(g|e) p(h|e, f) p(e|d, c) p(d) p(c|a, b) p(a) p(b). We're no longer tempted to *see* this formula as a structural representation (even if the pattern of statistical dependencies it expresses is the same). We're more naturally inclined to see *that* just like a series of "0"s and "1", or a series of electric discharges in my computer—something we have no reason to even remotely suspect is structurally similar to any target **T** one is trying to model.

[38] Or a series of ink marks when the article will be printed.

5 Conclusions: a dilemma for the cognitive neuroscience revolution.

Suppose my arguments are on the right track: NSRVs have *not* been observed and there is no easy way to avoid this conclusion. This is ill-news for defenders of the "cognitive neuroscience revolution": NSRs are absolutely central to their account (Sect. 1). So, the question now is: what could revolutionaries do to save their explanatory project? Not much, I fear.

They could try to substitute NSRs with a different type of representation. But this move is unpromising. According to a popular account, there are three basic representational kinds—*icons, symbols and indices* (c.f. Peirce 1931–1958; von Eckart 1996). Now, icons represent by similarity—so neural icons *just are* NSRs, and thus icons are clearly not an option. Symbols represent by stipulation—and so it is not clear if neural symbols can exist: surely no one has *stipulated* the content of our neurons. And even allowing stipulative or stipulation-like processes to take place in the brain (say, as the upshot of a neural signaling game, see Skyrms, 2010) the vehicles of neural symbols, being *arbitrary*, do not allow their content to play any causal role within neurocognitive mechanisms. Thus, symbolic representations have no place in mechanistic explanations. Lastly, indices represent in virtue of certain causal relation with their targets—they are indicators. Now, neural indicators surely exist, see (Sect. 3.1). Yet, it is far from clear they qualify as representations in any robust sense—they seem to function as mere causal mediators in our neurocognitive systems (Ramsey, 2003, 2007).

Should then the mechanistic approach to cognitive neuroscience be purged of representational commitments? Some claim this is the case (Kohar, 2023). This, however, would be an *extremely painful* revision of our current neuroscientific practices. Cognitive science is ripe with representational talk, and cognitive neuroscience is no exception. A non-representational mechanistic cognitive neuroscience would thus force us to revise and reinterpret a huge mass of experimental data. It would also force us to find a novel, non-representational lexicon with which to express and communicate the relevant cognitive-scientific findings. This surely is a tall order—one that proponents of the "cognitive neuroscience revolution" do not seem willing to execute.

The only way I see to avoid that non-representational revision, however, seems to be by foregoing one's realistic commitments to NSRs (or at least to NSRVs). The talk of neural maps and models, then, should not be interpreted as referring to real, neurally realized, map- and model- like structures. Rather, neural maps and models are just convenient linguistic tools to understand, track, or make sense of our neurocognitive activities (see Sprevak, 2013; Egan, 2020; Coelho Mollo, 2021; Cao, 2022 for similar views of representations).[39] But to adopt such a construal of NSRs or NSRVs amounts to abandoning one's mechanistic commitments, at least insofar mechanistic explanations are *ontic* explanations. But the commitment to mechanism is a core part of the "cognitive neuroscience revolution", and so abandoning it seems to abandon the "cognitive neuroscience revolution" project.

[39] See (Ramsey 2020) for acute criticism of some such accounts.

It seems, then, that defenders of NSRs face a dilemma: they either have to let go of their commitment to representationalism to keep their commitment to mechanistic explanations, or vice versa. The choice is theirs.

Acknowledgements Thanks to (in random order) Marco Viola, Davide Coraci, Jonny Lee and Sanja Sreckovic for having read and commented upon several previous poorly written and half-baked versions of this paper. Thanks to (again, in random order) Erik Thomson, Bryce Huebner and Carl Sachs for an extremely insightful exchange via Twitter on cortical maps and structural representations. This paper has also been presented at a number of conferences and workshops—in particular: the 4th International Conference in Philosophy of Mind in Braga (Portugal), the *Representational Penumbra* workshop held in Valencia (Spain), the British Society for Philosophy of Science conference held in Bristol (UK), the European Congress of Analytic Philosophy held in Vienna (Austria), the European Society of Philosophy and Psychology conference in Prague (Czech Republic), and the first online conference of the International Society for Philosophy and the Mind Sciences. I wish to thank the audience of all these conferences for their engaging questions and challenges. A special thanks goes to: Marc Artiga, Manolo Martinez, Peter Schulte, Nick Shea, Rosa Cao and Krys Dolega (again, in random order) for the several challenges they raised to the arguments I present here. I swear I will answer them all in a follow-up paper (and there will *really* be a follow-up paper, go read the *Appendix*)! A thanks also to the anonymous referees—with an apology for having forced them to sit through this gargantuan paper multiple times.

Author contributions MF is the sole author of the paper.

Funding This research was founded by the FWO grant "Towards a globally non-representational theory of the mind" (Grant Number 1202824N).

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The author declares no conflict of interests.

Appendix: on distinguishing types of structural representations (and why it matters)

During the review process, a reviewer (which I thank) has met many of the claims here with a number of reasonable observations on how structural representations are understood in the literature. And whilst (at least insofar this paper is concerned) the reviewer and I seem to have agreed to disagree, there is *something* to their observation—something that, I believe, points to the fact that, in the literature, the term "structural representation" is systematically ambiguous. Whilst this is not the place where to dispel this ambiguity,[40] I wish to point it out—if anything, to address a number of potential objections to my claim or misunderstandings of this paper.

Throughout the paper, I have relied on Gładziejewski's (2015, 2016) account of structural representations. Such an account is explicitly guided by the image of a cartographic map: a *single vehicle* whose constituents are enveloped in a web of relations

[40] I have an in-progress paper on this matter whose preprint can be consulted on my private website (https://marcofacchinmarcof.wixsite.com/site). Thanks to this anonymous referee for having motivated me to write it!

that mirror the web of relations of the constituents of a target, thereby making the former a representation of the latter. On such a—hopefully by now familiar—view of structural representations, the constituents of the whole vehicle are—in a way—representations too, whose representational status *derives* from the representational status of the whole vehicle (cf. Cummins, 1996). Given the popularity of Gładziejewski's (2015, 2016) account, and the fact that it is constantly referred to in the cognitive neuroscience revolution literature, it is reasonable to treat this as *the standard* understanding of structural representations (at least in that corner of philosophy). This is understanding of structural representation has been the target of my attack, and I won't comment any further on it—if not to notice two things: (a) the account spells out a specific "functional profile" for structural representations—telling us that they *function as representations by functioning as maps* (see Gładziejewski, 2015)[41]—and (b) that such an account is markedly *anti-csymbolic*. Thusly understood, structural representations *can't be* arbitrary symbols, for they *can't be arbitrary*: their very *physical shape* connects them to their targets (cf. Williams & Colling, 2017). Insofar "classic", rule-and-representation based cognitive science is symbolic, then, this account of structural representations is *anti-classical*.

As the reviewer correctly noticed, however, entities satisfying the description above are not the only referents philosophers grace with the title of structural representations. William Ramsey (2007) and, more recently, Matej Kohar (2023) used the term to refer to what I'll here call (for reasons that will soon be manifest) *carriers of structural contents*.[42] According to their usage, the term "structural representation" refers to individual vehicles belonging to a set of vehicles, the relations amongst which "mirror" the relations holding amongst the elements of some target domain. So, both according to my (and Gładziejewski's) usage and the structural content usage, the term "structural representation" refers to an individual vehicle. Yet, in my usage the structural similarity holds amongst an individual vehicle and its target, whereas in the structural content usage the similarity holds amongst the *set* each individual structural representation is part of, and some target domain. These are clearly different things.

Why call the entities satisfying the description above "carriers of structural contents"? Because what this account gives us is an account of *why* each individual vehicle of the set represents what it represents. Each vehicle represents what it represents because it is part of a set of vehicles, the relations amongst which make *the whole set* structurally similar to a target domain. Such a view of structural representations assigns a *content* to each vehicle based on its "place" in the overall similarity, but it remains utterly silent about its functional profile (which is left undefined) and their physical shape. Indeed, the vehicles carrying structural contents *can* be arbitrary—at least to the extent to which their arbitrary physical shapes do not interfere with them standing in the appropriate relations with each other.

Carriers of structural content can thus be coherently mashed with classical, symbolic, rules-and-representations based cognitive science. To see why this is the case, it

[41] This shouldn't be read as entailing that it spells out *only* the functional profile. Presumably, the content of such structural representations is in fact grounded in the similarity they bear to their targets.

[42] Carriers of structural contents surfaced in many places in the argument I developed in the paper, esp in (**§§ 3.1** and **3.3**). In all these cases, I argued that they are not structural representations in the relevant sense at play—that is, they don't satisfy Gładziejewski's account.

is sufficient to notice that Cummins's (1989) account of content for *classical* cognitive science is a particular incarnation of what I've been calling structural contents.[43] In the view Cummins originally proposed, computational states (the symbols of classical cognitive science) represent what they represent in virtue of the fact that the computational state transitions holding amongst them "mirror" certain relevant relations in a target domain. So, these vehicles represent what they represent *in virtue of* the fact that certain *computational relations* (mirroring the relevant relations of a target domain) hold among them. On some accounts, then, classical, symbolic representations can be structural contents—and can thus be called structural representations according to one usage of the term—which, however, it is not (and indeed cannot) be the relevant usage of the term made by defenders of the cognitive neuroscience revolution.

Similarly, *indicators* and *detectors* can qualify as carriers of structural contents—at least given the arguments offered by (Facchin, 2021b; Nirshberg & Shapiro, 2020).[44] On such views, individual indicators represent what they represent (and indicate what they indicate) in virtue of a specific structural similarity holding between the set of indicator states and the indicated target: indication is a special case of structural similarity (at least, if Facchin, Nirshberg and Shapiro are correct). Since—as argued in (Sects. 3.1 and 3.3) individual indicator states can't be constituents of a larger vehicle, we're seemingly forced to interpret them as individual vehicles of structural contents. So, indicators and detectors too can be said to be structural representations in one sense of the term, though not in the sense relevant to the cognitive neuroscience revolution.

Such a distinction between structural representations and carriers of structural contents, I believe, can be mobilized to *make sense* of why structural representations seem *both* to be everywhere *and* to systematically elude our gaze (as I argued above).

Consider first neuronal responses—both *individually* and *collectively* (as they are considered, for example, in representational similarity analysis, see Sect. 3.3) Individual neuronal responses are naturally classified as indicators (cf. Section 3.1), and so as carriers of structural contents (at least, if Facchin, Morgan, Shapiro and Nirshberg are on the right track). Sets of neuronal responses are also naturally read as carriers of structural contents—at least insofar the structural similarity holds between the *entire set of responses* and some target domain (cf. Section 3.5). So, whilst both are structural representations in some sense, they're not structural representations in the *relevant*, cognitive neuroscience revolution validating sense.

Consider now inner simulations and emulations. Such representations are often invoked in cognitive neuroscience (e.g. Csibra, 2008; Grush, 2004) and are taken as *bona fide* cases of structural representations. And indeed, they are carriers of structural contents: individual states of the simulation or emulation need not structurally resemble anything—only the *entire process* must. And since the process can't plausibly be considered an individual vehicle (cf Sects. 3.1 and 3.3), then we're left with carriers of

[43] And indeed, (Cummins 1989) is the account that Ramsey (2007) refers to when introducing structural representations *in the context of classic, rule and representation based theories of cognition*. For another example, see Kosslyn's (1983) "quasi-pictorial" representations.

[44] Though the two might be distinct. See the preprint I mentioned in footnote 40.

structural contents.[45] Again, simulations and emulations are structural representations *in some sense*, but that sense is not the one relevant for the cognitive neuroscience revolution. This, as the reviewer noticed, is a big problem for the cognitive neuroscience revolution. Arguably, their theoretical commitments make them unable to capitalize on (and are actually incompatible with, see below) the most widespread type of structural representation in the current neuroscientific literature.

Consider lastly the fact that I've hunted for structural representation roughly at the *implementation* level, looking at the actual neural machinery (allegedly) doing the representing. Can't structural representations be found at higher, roughly algorithmic, levels of abstraction? Yes, but only in the sense that *carriers of structural contents* can be found at such levels of abstraction.[46] For, in this case, the *physical shape* of the vehicles is not relevant to their being structural representations (i.e. carriers of structural contents)—only their relations are. In contrast, in the case of structural representations in the relevant sense, the physical shape of the vehicles is *essential* to their status as a structural representation. Their implementation matters for their representational state. Hence, they should be found at the implementation level.

The distinction between structural representations in the relevant sense and carriers of structural contents, then, allows us to make sense of *both* the seemingly omnipresence of structural representations (indeed, carriers of structural contents appear to be widespread) *and* their actual disappearance on closer inspection (nothing seems to satisfy Gładziejewski's account). A natural question, at this point, is whether the cognitive neuroscience revolution may ditch Gładziejewski's structural representations in favor of carriers of structural contents. The answer, I think, is negative. For, carriers of structural contents are entirely compatible with classic cognitive science. By adopting them, the cognitive neuroscience revolution would stop being a *revolution*. Worse still, the contents carried by carriers of structural contents is independent from their vehicle properties. So, it can't play the relevant causal role played by the content of structural representations (Sects. 1 and 2). As such, the contents of carriers of structural contents are not explanatory assets defenders of the cognitive neuroscience revolution can count upon.

Does this mean that structural representations, in the relevant sense discussed here, will *never* be observed? Not necessarily. Perhaps, as the reviewer suggests, we might be able to observe them thanks to a *methodological shift*—diverting our attention from neuronal responses (which, at best, carry structural contents) from spontaneous, endogenous and "decoupled", non-stimulus-driven neural activity. Whilst such a shift in attention faces some methodological challenges (see Sect. 3), it might be possible to face them, and observe structural representations in the relevant sense.

Even in this case, however, neural structural representation (in the relevant sense) would remain *unobserved*—they may populate our brains, but we have not seen them yet. What we're left with, then, are some thorny issues for the defenders of the cognitive neuroscience revolution to solve, together with the need to disentangle various distinct

[45] At least unless defenders of the cognitive neuroscience revolution are willing to significantly modify and complexify the mechanistic metaphysics grounding their view, allowing for non-synchronic constitutive relations.

[46] Emulators and inner simulations may be one such case.

senses of the term "structural representations". And the latter is definitely a task for a different paper.

References

Aflalo, T. N., & Graziano, M. S. (2006). Partial tuning of motor cortex neurons to final posture in a free-moving paradigm. *Proceedings of the National Academy of Sciences, 103*(8), 2909–2914.

Albers, A. M., et al. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427–1431.

Anderson, M. L. (2014). *After phrenology*. The MIT Press.

Anderson, M. L., & Champion, H. (2022). Some dilemmas for an account of neural representation: A reply to Poldrack. *Synthese, 200*(2), 169.

Artiga, M., & Sebastián, M. A. (2018). Informational theories of content and mental representation. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-018-0408-1

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience, 22*(6), 359–371.

Backer, B., et al. (2022). Three aspects of representation in neuroscience. *Trends in Cognitive Sciences, 26*(11), 942–958.

Baumgartner, M., Casini, L., & Krickel, B. (2020). Horizontal surgicality and mechanistic constitution. *Erkenntnis, 85*(3), 417–430. https://doi.org/10.1007/s10670-018-0033-5

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

Bechtel, W. (2014). Investigating neural representations: The tale of place cells. *Synthese, 193*, 1287–1321.

Bielecka, K., & Miłkowski, M. (2020). Error detection and representational mechanisms. In J. Smortchkova, K. Dolega, & T. Schicht (Eds.), *What are mental representations?* (pp. 287–317). Oxford University Press.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2022). A connectivity-constrained computational account of topographic organization in primate high-level visual cortex. *Proceedings of the National Academy of Sciences, 119*(3), e2112566119.

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese, 193*(5), 1509–1534.

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience, 28*, 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

Brette, R. (2015). Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Frontiers in Systems Neuroscience, 9*, 151.

Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences, 42*, e215.

Bruineberg, J., & Rietveld, E. (2019). What's inside your head once you've figured out what your head's inside of. *Ecological Psychology, 31*(3), 198–217.

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology, 81*, 55–79.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. *Biology & Philosophy, 31*, 527–550.

Buzsaki, G. (2006). *Rhythms in the brain*. Oxford University Press.

Cao, R. (2022). Putting representations to use. *Synthese, 200*(2), 151.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience, 26*(1), 132–142.

Chakrabarty, S., & Martin, J. H. (2000). Postnatal development of the motor representation in primary motor cortex. *Journal of Neurophysiology, 84*(5), 2582–2594.

Chemero, A. (2009). *Radical embodied cognitive science*. The MIT Press.

Churchland, P. M. (1992). *A neurocomputational perspective*. The MIT Press.

Churchland, P. M. (1995). *The engine of reason, the sit of the soul*. The MIT Press.

Clark, A. (1997). *Being there*. The MIT Press.

Coelho Mollo, D. (2021). Deflationary realism: Representation and idealisation in cognitive science. *Mind & Language, 37*(5), 1048–1066.

Connolly, A. C., et al. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience, 32*(8), 2608–2618.

Coraci, D. (2022). Representations and processes: What role for multivariate methods in cognitive neuroscience? *Rivista Internazionale Di Filosofia e Psicologia, 13*(3), 187–199.
Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.
Csibra, G. (2008). Action mirroring and action understanding: An alternative account. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition* (pp. 435–459). Oxford University Press.
Cummins, R. (1989). *Meaning and mental representation*. MIT Press.
Cummins, R. (1996). *Representations, targets, attitudes*. The MIT Press.
Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. The MIT Press.
Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences, 1296*(1), 108–134.
Dayan, P., & Abbott, L. F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT Press.
De Angelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience, 19*(4), 1398–1415.
Dennett, D. C. (1996). *Darwin's dangerous idea*. Penguin.
de Wit, M. M., & Matheson, H. E. (2022). Context-sensitive computational mechanistic explanation in cognitive neuroscience. *Frontiers in Psychology, 13*, 903960.
Downey, A. (2018). Predictive processing and the representation wars: A victory for the eliminativist (via fictionalism). *Synthese, 195*, 5115–5139.
Dretske, F. (1988). *Explaining behavior*. The MIT Press.
Egan, F. (2020). A deflationary account of mental representations. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* (pp. 26–54). Oxford University Press.
Facchin, M. (2021a). Predictive processing and anti-representationalism. *Synthese, 199*(3–4), 11609–11604.
Facchin, M. (2021b). Structural representations do not meet the job description challenge. *Synthese, 199*(3), 5479–5508.
Favela, L. H., & Machery, E. (2023). Investigating the concept of representation in the neural and psychological sciences. *Frontiers in Psychology, 14*, 1165622.
Fodor, J. A. (1981). The mind-body problem. Scientific American 244 (January 1981). Reprinted in J. Heil, (Ed.) (2004a), Philosophy of Mind: A Guide and Anthology (168–82). Oxford University Press
Frisby, S. L., et al. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences., 27*(3), 258–281.
Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences., 360*(1456), 815–836.
Gessell, B., Geib, B., & De Brigard, F. (2021). Multivariate pattern analysis and the search for neural representations. *Synthese, 199*(5–6), 12869–12889.
Gładziejewski, P. (2015). Explaining cognitive phenomena with internal representations: A mechanistic perspective. *Studies in Logic, Grammar and Rhetoric, 40*(1), 63–90.
Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese, 193*(2), 559–582.
Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: Causally relevant and distinct from detectors. *Biology and Philosophy, 32*(3), 337–355.
Gordon, E. M., et al. (2022). A mind-body interface alternates with effector-specific regions in motor cortex. *Nature*. https://doi.org/10.1038/s41586-023-05964-2
Graziano, M. S. (2011). Cables vs. networks: old and new views on the function of motor cortex. *The Journal of Physiology, 589*(Pt 10), 2439.
Graziano, M. S. (2016). Ethological action maps: A paradigm shift for the motor cortex. *Trends in Cognitive Sciences, 20*(2), 121–132.
Graziano, M. S., & Aflalo, T. N. (2007). Mapping behavioral repertoire onto the cortex. *Neuron, 56*(2), 239–251.
Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*(3), 377–396.
Grush, R., & Mandik, P. (2002). Representational parts. *Phenomenology and the Cognitive Sciences, 1*(3), 389–394.
Ha, D., & Schmidhuber, J. (2018a). Recurrent world models facilitate policy evolution. In S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems 31* (pp. 2451–2463). Curran Associates.

Ha, D., Schmidhuber, J. (2018b). World models. Preprint. ArXiv:18.0310122.
Hartmann, K., et al. (2016). Embedding a panoramic representation of infrared light in the adult rat somatosensory cortex through a sensory neuroprosthesis. *Journal of Neuroscience, 36*(8), 2406–2424.
Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for motor generation. In T. Ono, G. Matsumoto, R. R. Llinas, A. Bethoz, R. Norgren, H. Nishijo, R. Tamura (Eds.), *Excepta medica international congress system* (Vol. 1250), (pp. 575–590). Elsevier.
Haueis, P. (2018). Beyond cognitive myopia: A patchwork approach to the concept of neural function. *Synthese, 195*(12), 5373–5402.
Haxby, J., et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425–2430.
Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience, 37*, 435–456.
Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*(1), 215–243.
Hurley, S. (1998). *Consciousness in action*. Cambridge University Press.
Hutto, D., & Myin, E. (2013). *Radicalizing enactivism*. The MIT Press.
Illari, P. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis, 78*, 237–255.
Isaac, A. M. (2013). Objective similarity and mental representation. *Australasian Journal of Philosophy, 91*(4), 683–704.
Itskov, P. M., et al. (2011). Hippocampal representation of touch-guided behavior in rats: Persistent and independent traces of stimulus and reward location. *PLoS ONE, 6*, e16462. https://doi.org/10.1371/journal.pone.0016462
Johnson-Laird, P. (1983). *Mental models*. Harvard University Press.
Kaplan, H. S., & Zimmer, M. (2020). Brain-wide representations of ongoing behavior: A universal principle? *Current Opinion in Neurobiology, 64*, 60–69.
Kelso, S. (1995). *Dynamic patterns*. The MIT Press.
Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing, 8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2
Kirchhoff, M. (2014). Extended cognition & constitution: Re-evaluating the constitutive claim of extended cognition. *Philosophical Psychology, 27*(2), 258–283.
Kirchhoff, M. D. (2015). Extended cognition & the causal-constitutive fallacy: In search for a diachronic and dynamical conception of constitution. *Philosophy and Phenomenological Research, 90*(2), 320–360.
Kirchhoff, M. D., & Kiverstein, J. (2021). Diachronic constitution. *Preprint*. http://philsci-archive.pitt.edu/19690/
Kiverstein, J., & Kirchhoff, M. D. (2023). Dissolving the causal-constitution fallacy: Diachronic constitution and the metaphysics of extended cognition. In M. O. Caspar & G. F. Artese (Eds.), *Situated cognition research: Methodological foundations.* Springer.
Kohar, M. (2023). *Neural machines: A defense of non-representationalism in cognitive neuroscience*. Springer.
Kohler, E., et al. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*(5582), 846–848.
Kosslyn, S. (1983). *Ghosts in the mind's machine*. W.W. Norton.
Kraus, B. J., Robinson, R. J., White, J. A., Eichenbaum, H., & Hasselmo, M. E. (2013). Hippocampal "time cells": Time versus path integration. *Neuron, 78*(6), 1090–1101.
Krickel, B. (2023). Extended cognition and the search for the mark of constitution—a promising strategy? In M. O. Casper & G. F. Artese (Eds.), *Situated Cognition Research - Methodological foundations.* Springer.
Kriegeskorte, N., et al. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008
Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences, 17*(8), 401–412.
Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the onion of brain representations. *Annual Review of Neuroscience, 42*, 407–432.
Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A, 68*, 58–67.
Kwan, H. C., et al. (1978). Spatial organization of precentral cortex in awake primates. *II. Motor Outputs. Journal of Neurophysiology, 41*(5), 1120–1131.

Lee, J. (2019). Structural representations and the two problems of content. *Mind & Language, 34*(5), 606–626.
Lee, J. (2021). Rise of the swamp creatures. *Philosophical Psychology, 34*(6), 805–828.
Lee, J., & Calder, D. (2023). The many problems with S-representation (and how to solve them). *Philosophy and the Mind Sciences*. https://doi.org/10.33735/phimisci.2023.9758
Lee, A. Y., et al. (2022). The structure of analog representation. *Noûs, 2022*, 1–28. https://doi.org/10.1111/nous.12404
Leuridan, B., & Lodewyckx, T. (2021). Diachronic causal constitutive relations. *Synthese, 198*, 9035–9065.
Maley, C. (2021a). Analog computation and representation. *The British Journal of Philosophy of Science*. https://doi.org/10.1086/715031
Maley, C. J. (2021b). The physicality of representation. *Synthese, 199*(5–6), 14725–14750.
Maley, C. (2023). Icons, magnitudes and their parts. Forthcoming in Critica: Revista Hispanoamericana de Filosofia.
Martin, J. H., et al. (2000). Impairments in prehension produced by early postnatal sensorimotor cortex activity blockade. *Journal of Neurophysiology, 83*, 895–906.
Martin, J. H., et al. (2005). Effect of forelimb use on postnatal development of the forelimb motor representation in primary motor cortex of the cat. *Journal of Neurophysiology, 93*(5), 2822–2831.
Martinez, M., & Artiga, M. (2021). Neural oscillations as representations. *The British Journal of Philosophy of Science*. https://doi.org/10.1086/714914
Mataric, M. (1991). Navigating with a rat's brain: A neurobiologically inspired model for robot spatial representation. In J. A. Meyer & S. Wilson (Eds.), *From animals to animats 1* (pp. 169–175). The MIT Press.
McClelland, J. L., et al. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). The MIT Press.
McLendon, H. J. (1955). Uses of similarity of structure in contemporary philosophy. *Mind, 64*(253), 79–95.
McNamee, D., & Wolpert, D. M. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems, 2*, 339–364.
Mesulam, M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Annals of Neurology, 64*(4), 367–378.
Morgan, A. (2014). Representations gone mental. *Synthese, 191*(2), 213–244.
Morgan, A., & Piccinini, G. (2018). Towards a cognitive neuroscience of intentionality. *Minds and Machines, 28*, 119–139.
Moser, E. I., Kropff, E., & Moser, M. B. (2008). Place cells, grid cells, and the brain spatiotemporal representation system. *Annual Review Neuroscience, 31*, 69–89. https://doi.org/10.1146/annurev.neuro.31061307.090723
Neander, K. (2017). *A mark of the mental*. The MIT Press.
Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science, 313*(5792), 1431–1435.
Nirshberg, G. (2023). Structural resemblance and the causal role of content. *Erkenntnis*, 1–20.
Nirshberg, G., & Shapiro, L. (2020). Structural and Indicator representations: A difference in degree, not in kind. *Synthese*. https://doi.org/10.1007/s11229-020-02537-y
O'Brien, G. (2015). How does mind matter? Solving the content causation problem. In T. K. Metzinger & J. M. Windt (Eds.), *Open mind*. Mind Group. https://doi.org/10.15502/9783958570146
O'Brien, G., & Opie, J. (2009). The role of representation in computation. *Cognitive Processing, 10*, 53–62.
O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Clarendon Press.
O'Regan, K. (2011). *Why doesn't red sound like a bell*. Oxford University Press.
Orlandi, N. (2020). Representing as coordinating with absence. In J. Smortchkova, K. Dołega, & T. Schlicht (Eds.), *What are mental representations?* (pp. 101–134). Oxford University Press.
Peirce, C. S. (1931–1958). Collected papers of Charles Sanders Peirce. In: P. Hartshorne, P. Weiss, & A. Burks (Eds.) (Vols. 1–8). Harvard University Press
Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain, 60*(4), 389–443.
Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man; a clinical study of localization of function*. Macmillan.
Pezzulo, G. (2008). Coordinating with the future: The anticipatory nature of representation. *Minds and Machines, 18*, 179–225.
Piccinini, G. (2015). *Physical computation*. Oxford University Press.

Piccinini, G. (2020a). *Neurocognitive mechanisms*. Oxford University Press.
Piccinini, G. (2020). Nonnatural mental representations. In G. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* Oxford University Press.
Piccinini, G. (2022). Situated neural representations: solving the problems of content. *Frontiers in Neurorobotics*. https://doi.org/10.3389/fnbot.2022.846979
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese, 183*, 283–311.
Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences, 18*(9), 451–456.
Poldrack, R. (2020). The physics of representation. *Synthese, 199*, 1307–1325.
Quiroga, R., et al. (2005). Invariant visual representation by single neurons in the human brain. *Nature, 435*(7045), 1102–1107.
Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433–447.
Ramsey, W. (2003). Are receptors representations? *Journal of Experimental & Theoretical Artificial Intelligence, 15*(2), 125–141.
Ramsey, W. (2007). *Representation reconsidered*. Cambridge University Press.
Ramsey, W. (2016). Untangling two questions about mental representation. *New Ideas in Psychology, 40*, 3–12.
Ramsey, W. (2020). defending representation realism. In J. Smortchkova, K. Dolega, & T. Schlicht (Eds.), *What are mental representations?* (pp. 54–84). Oxford University Press.
Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging object representations in the visual system predict reaction times for categorization. *PLOS Computational Biology, 11*(6), e1004316.
Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axx023
Rizzolatti, G., & Sinigaglia, C. (2023). *Mirroring brains*. Oxford University Press.
Rosche, W., & Sober, E. (2019). Disjunction and distality: The hard problem for purely probabilistic causal theories of mental content. *Synthese*. https://doi.org/10.1007/s11229-019-02516-y
Roskies, A. L. (2021). Representational similarity analysis in neuroimaging: Proxy vehicles and provisional representations. *Synthese, 199*(3–4), 5917–5935.
Rutar, D., Wiese, W., & Kwisthout, J. (2022). From representations in predictive processing to degrees of representational features. *Minds and Machines, 32*(3), 461–484.
Schieber, M. H. (2001). Constraints on somatotopic organization in the primary motor cortex. *Journal of Neurophysiology, 86*(5), 2125–2143.
Segundo-Ortin, M., & Hutto, D. D. (2021). Similarity-based cognition: Radical enactivism meets cognitive neuroscience. *Synthese, 198*(Suppl 1), 5–23.
Seth, A. K. (2015). The cybernetic bayesian brain. In T. Metzinger, J. Windt (Eds.), *Open* MIND. The MIND Group. https://doi.org/10.15502/9783958570108
Shagrir, O. (2012). Structural representations and the brain. *The British Journal for the Philosophy of Science., 63*(3), 519–545.
Shagrir, O. (2018). The brain as an input–output model of the world. *Minds and Machines, 28*, 53–75.
Shea, N. (2018). *Representation in Cognitive Science*. Oxford University Press.
Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science, 80*(5), 958–970.
Silberstein, M. (2021). Constraints on localization and decomposition as explanatory strategies in the biological sciences 20. In M. Viola & F. Calzavarini (Eds.), *Neural Mechanisms: new challenges in the philosophy of neuroscience.* Springer.
Skyrms, B. (2010). *Signals: Evolution, learning, and information*. OUP.
Sprevak, M. (2013). Fictionalism about neural representations. *The Monist, 96*(4), 539–560.
Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT Press.
Sun, C., Yang, W., Martin, J., & Tonegawa, S. (2020). Hippocampal neurons represent events as transferable units of experience. *Nature Neuroscience, 23*(5), 651–663.
Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese, 87*(3), 449–508.
Tani, J. (2007). On the interactions between top-down anticipation and bottom-up regression. *Frontiers in Neurorobotics, 1*, 2.
Tani, J. (2016). *Exploring robotic minds*. Oxford University Press.
Thomson, E., & Piccinini, G. (2018). Neural representations observed. *Minds and Machines, 28*, 191–235.

Tootell, R. B., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic Organization. *Journal of Neuroscience, 8*(5), 1531–1568.
Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLoS Comput Biol, 16*(4), e1007805.
Van Bree, S. (2023). A critical perspective towards mechanisms in cognitive neuroscience: Towards unification. *Perspectives on Psychological Sciences*. https://doi.org/10.1177/17456916231191744
Van der Weel, F. R., Sokolovskis, I., Raja, V., & van der Meer, A. L. (2022). Neural aspects of prospective control through resonating taus in an interceptive timing task. *Brain Sciences, 12*(12), 1737.
Van Gelder, T. (1991). What is the "D" in "PDP"? A survey of the concept of distribution. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Philosophy and connectionist theory.* Routledge.
Vilarroya, O. (2017). Neural representation. A survey-based analysis of the notion. *Frontiers in Psychology, 8*, 1458.
Von Eckardt, B. (1996). *What is cognitive science?* The MIT Press.
Wassermann, E. M., et al. (1992). Noninvasive mapping of muscle representations in human motor cortex. *Electroencephalography and Clinical Neurophysiology/evoked Potentials Section, 85*(1), 1–8.
Westlin, C., et al. (2023). Improving the study of brain-behavior relationships by revisiting basic assumptions. *Trends in Cognitive Sciences., 27*(3), 246–257.
Wiese, W. (2016). What are the contents of representations in predictive processing? *Phenomenology and the Cognitive Sciences, 16*, 715–736.
Wiese, W. (2017). *Experienced wholeness*. The MIT Press.
Williams, D. (2017). Predictive processing and the representation wars. *Minds and Machines, 28*(1), 141–172.
Williams, D., & Colling, L. (2017). From symbols to icons: The return of resemblance in the cognitive science revolution. *Synthese, 195*(5), 1941–1967.
Wood, E. R., et al. (1999). The global record of memory in hippocampal neuronal activity. *Nature, 397*(6720), 613–616.
Wood, E. R., et al. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron, 27*(3), 623–633.
Woolsey, et al. (1952). Patterns of localization in precentral and" supplementary" motor areas and their relation to the concept of a premotor area. *Research Publications-Association for Research in Nervous and Mental Disease, 30*, 238–264.
Yarrow, S., et al. (2014). Detecting and quantifying topography in neural maps. *PLoS ONE, 9*(2), e87178.