

Neurophenomenal structuralism and the role of computational context

Marlo Paßler^{1, 2, 3, *} 

Adrien Doerig^{4, 5} 

Special Issue: *Structuralism in the Science of Consciousness*, edited by Sascha Benjamin Fink & Andrew Lee

¹ Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Germany ² RTG 2386 “Extrospection”, Humboldt-Universität zu Berlin, Germany ³ Otto-von-Guericke University, Institute of Philosophy, Magdeburg, Germany ⁴ Department of Education and Psychology, Freie Universität Berlin, Germany ⁵ Bernstein Center for Computational Neuroscience, Berlin, Germany

* Primary contact: marlopassler@hu-berlin.de

Paßler, M., & Doerig, A. (2025). Neurophenomenal structuralism and the role of computational context. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11824>

Abstract

Neurophenomenal structuralism posits that conscious experiences are defined relationally and that their phenomenal structures are mirrored by neural structures. While this approach offers a promising framework for identifying neural correlates of the contents of consciousness, we argue that merely establishing structural correspondences between neural and phenomenal structures is insufficient. This paper emphasizes the critical role of computational context – the network of neural processes within which a given neural activation pattern is used – in determining content. We introduce four criteria to evaluate if neural structures are viable candidates for neural correlates of contents of consciousness within this framework. These criteria highlight that, for neural structures to mirror phenomenal structures meaningfully, they must be actively exploited by the brain’s downstream processes and influence behavior in a structure-preserving way. Our analysis demonstrates that purely anatomical and overly exhaustive causal structures fail to meet specific criteria, whereas activation structures can succeed, provided they are embedded within the appropriate computational context. Our findings challenge local structuralist theories, which overlook the content-constituting role of computational context. We conclude that incorporating computational context is essential for any structuralist account of consciousness.

Keywords: Computational Context • Consciousness • Criteria • Representation • Structuralism



1 Introduction

In the ever-evolving field of consciousness science, an approach known as ‘neurophenomenal structuralism’ (NPS) is making waves. Championed by scholars such as Fink et al. (2021), Kob (2023), Lau et al. (2022), Lyre (2022), and Tsuchiya & Saigo (2021), NPS offers a fresh perspective on the relation between neural states and conscious experiences. Central to this approach is the assertion that one cannot characterize a conscious experience in isolation. For example, we cannot characterize the experience of blue. However, we can characterize the relationships between experiences. Roughly, blue is perceptually more similar to purple than to orange. Further, neurophenomenal structuralists claim that this phenomenal structure is mirrored in neural structure. Simply put, they propose that the neural vehicles of blue percepts are more similar to the neural vehicles of purple percepts than to those of orange ones. As such, experiences are relationally defined, and the resulting similarity structures are systematically mirrored by their neural vehicles. This mapping is envisioned by these authors as an isomorphic, structure-preserving correspondence, pinpointing the neural correlates of consciousness (NCC), or at least its content-constituting parts, the neural correlates of the contents of consciousness (NCCC).

This paper outlines criteria for identifying which types of neural structure are plausible candidates for mirroring phenomenal structure, as suggested by NPS. In other words, assuming that NPS is true, which kinds of neural structures should we look for? Our argument builds upon philosophical literature on structural representation, emphasizing that a structure counts as representational only if it is actively exploited by the system that embodies it (Gładziejewski and Miłkowski, 2017; Millikan, 1989, 2004; Shea, 2014, 2018). Adapting the idea of an *exploitable isomorphism*, we argue that a meaningful neurophenomenal mapping requires not just structural resemblance, but also evidence that the brain uses this neural structure *as structure*.

A simple example clarifies our point (Figure 1). Suppose the brain encodes colors in a neural map such that similar colors evoke overlapping activation patterns across a population of neurons: blue and purple activate

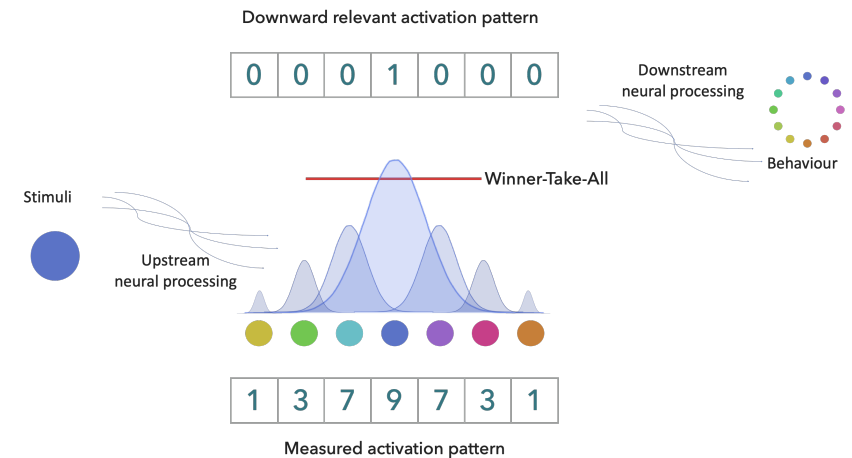


Figure 1: Illustration of ‘Winner-Take-All’ read-out: The figure shows a color-coding neural map reacting to a blue stimulus. Downstream neural processing employs a ‘winner-takes-all’ strategy, filtering out all but the strongest signal and reducing the rich, overlapping activation pattern to a simple binary signal. In this process, the nuanced relational information about color similarities in the measured activation pattern is lost.

more overlapping sets than blue and orange. Now imagine that a downstream process employs a “winner-takes-all” strategy: it selects only the single neuron with the strongest response, ignoring all other co-activated neurons. This read-out method collapses a rich, distributed activation pattern into a single discrete signal. Consequently, the nuanced similarity relations initially present (e.g., blue \approx purple \neq orange) are lost: orange, red, and blue are now encoded by distinct “winning” neurons, rendering them equidistant in neural terms. This illustrates the importance of computational context—specifically, how activation patterns are read out and exploited by downstream systems. If the relational information embedded in those patterns is ignored, any apparent correspondence between neural and phenomenal structure may be merely coincidental: an artifact of statistical correlation, disambiguation operations, or noise, rather than genuine structural encoding.

This highlights that even when neural activity patterns exhibit isomorphism, this does not necessarily imply that the brain itself uses this mapping in a structure-preserving or functionally meaningful way (Brette, 2018; Klein, 2010; Poldrack, 2006; Ritchie et al., 2019). For a neural structure to be a valid NCCC within NPS, it is not enough to just have a structural correspondence—the brain’s downward processes must actively use this information to shape subjective behavior.

From such considerations, we derive four criteria that any viable candidate NCCC must satisfy within NPS: Sensitivity, Organization, Exploitation, and Contextualization. Crucially, we emphasize that incorporating computational context is essential for any structuralist account of consciousness, as context dictates the interpretation of dimensions within neural activation spaces and, consequently, shapes the content of conscious experiences. This leads us to reject the notion that local relational configurations *alone* can sufficiently explain conscious content. By underscoring the importance of computational context, our approach refines and extends the structuralist program, demonstrating how its core claims can be empirically tested.

2 How to map out phenomenal structure

At the heart of NPS is the notion that experiences are best understood through their relationships with one another. Rather than describing a phenomenal quality—such as seeing red—in isolation, we characterize it by how it relates to other similar and dissimilar experiences. For example, red is perceptually similar to orange and pink, but dissimilar to blue and green. Such comparative judgments provide practical means to communicate what an experience feels like, and from them we can systematically construct “quality spaces” that capture these perceived similarities and differences (Clark and Clark, 1993; Rosenthal, 2010, 2015).

The construction of such quality spaces relies on psychophysical experiments, grounded in the operational insight that mental qualities are accessible “not only by way of consciousness, but also by way of their intimate and arguably essential role in perception” (Rosenthal, 2010). Participants

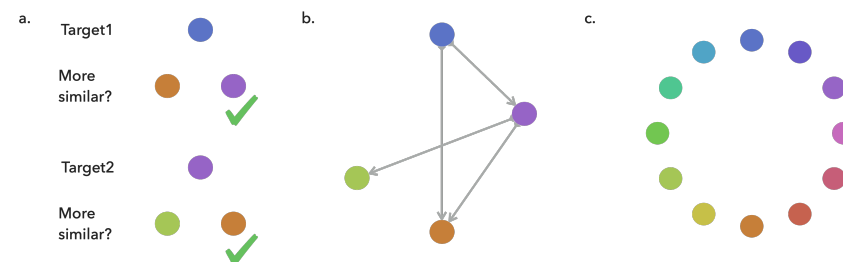


Figure 2: From Behavior to Quality Spaces: a.) Participants provide color-similarity reports, for example, by reporting which of two colors is more similar to a target color. b.) These similarity reports are translated into spatial distances, with colors judged to be more similar placed closer together. c.) Conducting these steps over a large number of stimuli allows researchers to construct quality spaces. In the case of colors, the quality space is approximately circular.

engage in tasks that capture perceived similarities between stimuli—such as indicating whether purple or orange is more similar to a reference color like blue (see Figure 2). These judgments are then converted into distances in an abstract space using statistical techniques such as multidimensional scaling (Mead, 1992). In these spaces, stimuli perceived as more alike are located closer together, while dissimilar ones are placed further apart. When carefully applied across a comprehensive stimulus set—such as the full range of visible colors—this method yields *quality spaces* that robustly reflect perceived similarities (Churchland, 2005; Kawakita et al., 2025; Palmer, 1999).

Over decades, psychophysics has refined its methods to probe the phenomenal structure of subjective experience more directly. A wide range of techniques—such as direct pairwise similarity judgments (Yoshida, 1964), multidimensional scaling (Mead, 1992), inverse multidimensional scaling (Kriegeskorte and Mur, 2012), and feature-weighted representational similarity analysis (Kaniuth & Hebart, 2022)—have been developed. In addition, the integration of machine learning methods has enhanced both the precision and scalability of quality space construction (Kawakita et al., 2025; Lötsch et al., 2019), yielding increasingly refined and robust estimates of

phenomenal structure (see also Kob, 2023).

This methodological stance differs from traditional NCC research, which typically relies on near-threshold stimuli to contrast conscious and unconscious processing (Baars, 1986). This contrastive approach studies why some states are conscious while others are not (what “turns on the light” of consciousness?). This is sometimes called the “consciousness question”; NPS focuses instead on the phenomenal character, or content, of conscious experiences (“what does the light reveal?”), sometimes called the “quality question” (Michel, 2024).

These two approaches are complementary, while it is possible that a common set of neural structures may inform both conscious and unconscious processing. For example, continuous priming paradigms have repeatedly shown that reaction times vary systematically with perceived similarity, with similar primes eliciting faster responses than dissimilar ones. This effect has been reliably demonstrated across diverse domains, including numbers (Hesselmann et al., 2015; Koechlin et al., 1999), faces (de Gardelle et al., 2011), orientations (Wilson et al., 2010), spatial locations (Vorberg et al., 2004), and object categories (Cohen et al., 2015; Haberkamp et al., 2019). The fact that the same neural structures may support conscious and unconscious behavior is not problematic for our purposes here, since NPS aims to isolate the NCCC — the specific neural structures that constitute what is consciously experienced (“quality question”).

A last point of discussion worth noting before diving into our main argument concerns the type of data on which quality spaces are based: first-person subjective data or third-person objective data? Put differently, when constructing a quality space, are behavioral reports merely third-person, objective *proxies* for phenomenal structure (an “indirect” view)? Or can quality spaces be taken as grounded *directly* in first-person subjective data (a “direct” view)? For our purposes, it is sufficient to note that behavioral data play a critical role in charting phenomenal structure from both perspectives. Even proponents of explicitly “first-person” methodologies acknowledge this reliance. As Tsuchiya et al. (2019) emphasize, “In consciousness science we rely on the link—known to each of us by first-hand experience—between conscious experience and behavioral reports. This link motivates

our collection of physical measurements including behavioral reports, and motivates theories about consciousness.”

In the remainder of this article, we thus set aside these questions to focus on our main points: First, does a mere structural match between neural activation patterns and phenomenal structure suffice to identify the neural substrates that determine the content of consciousness? We argue that it does not. Second, given this, what further criteria must candidate neural structures satisfy—beyond structural correspondence—to count as genuine NCCCs

3 Beyond correspondence: The importance of exploitability in identifying NCCCs

While NPS emphasizes a systematic mapping between neural and phenomenal structures, the precise nature of this systematicity remains underspecified. We propose that it should be understood in terms of exploitability: the neural structures that mirror phenomenal structure must also be the very structures used to generate the similarity judgments from which their corresponding quality space is derived. This requirement is not arbitrary, but we argue that it is already implicit in the NPS framework. Without it, the notion of a structural “mirror” risks collapsing into a coincidental resemblance with no real explanatory force. Our central claim is thus: **exploitability is essential for distinguishing genuine mirroring from mere resemblance**. The argument unfolds across four premises:

Premise 1 – neuro-phenomenal structural mapping: There is a structural mapping between neural and phenomenal structures.

Premise 2 – measurability assumption: Under standard conditions, quality spaces provide reliable approximations of phenomenal structure.

Premise 3 – empirical access to phenomenal structure: Quality spaces themselves are constructed from behavioral data (direct or indirect).

Consequently, behavior mediates our empirical access to phenomenal structure.

Premise 4 – requirement of structure-preservation: For this empirical access to be valid, there must also be a structure-preserving mapping between the neural structures that mirror phenomenal structures and the behaviorally measured quality spaces. Without such a mapping, phenomenal and behavioral structures could systematically come apart, undermining Premise 2.

Conclusion – neuro-behavioral structural mapping: Therefore, for NPS to be empirically tractable, the neural structures that mirror phenomenal structures must be able to impact behavior in a structure-preserving way.

Premise 1 is the basic assumption of NPS. Questions arise on both sides of the neurophenomenal mapping assumption. Some researchers debate whether perception and its neural vehicles are continuous and structural or instead rely on a more discrete, perhaps symbolic format (Herzog et al., 2020; Quilty-Dunn et al., 2022; Tee and Taylor, 2020; VanRullen and Koch, 2003). Others question whether conscious experience is as richly structured as it seems (Cohen et al., 2016; Kammerer, 2021; Kouider et al., 2007, 2010). We set aside these and other potential objections (Fink and Kob, 2023; Kleiner, 2024) to focus on the implications of accepting Premise 1.

Premise 2 builds on the continuous methodological development of psychophysics and quality space theory, as outlined in Chapter 2. Even if consciousness cannot be measured directly (see Feest, 2014; Irvine, 2012, 2013; Overgaard, 2015; Pauen and Haynes, 2021; Spener, 2022), behavioral measures such as similarity ratings can serve as reliable tools for indirectly approximating phenomenal structure, this is not to claim that behavior always tracks experience—coercion, biases, or attentional constraints may introduce distortions. However, under normal conditions, subjective reports are generally regarded as reliable, even by skeptics (Paßler, 2023). This premise ensures quality spaces successfully approximate phenomenal structure.

Premise 3 states that our empirical access to phenomenal structure via quality spaces is grounded in behavioral data (direct or indirect). Even if future techniques may enable us to infer aspects of phenomenal structure directly from neural activity, such inferences would still require calibration, validation, and constraint by behavioral reports. This grounding assumption is a cornerstone of consciousness science and is shared even by advocates of first-person methodologies. Any empirical theory of consciousness must preserve the claim that behavior carries informative links to experience; without this foundational link, no method—neural or otherwise—could be trusted to measure phenomenal structure at all.

Premise 4 states that given a neural structure mirrors phenomenal structure, it must be exploitable in a structure-preserving way to generate behavior that can be used to build accurate quality spaces. Indeed, the measurability assumption of Premise 2 supports the neurophenomenal mapping hypothesis of Premise 1 if and only if the mirroring neural structures can shape subjective behavior in a manner that preserves its structure. Without this causal link, quality spaces are decoupled from the neural structures that NPS claims mirror experience, undermining premise 2.

We conclude that, for NPS to remain both empirically tractable and theoretically viable, the neural structures that mirror phenomenal structure must also influence behavior in a structure-preserving manner. Put more precisely, the relevant neural structures must shape behavioral structure *in virtue of* their own structure. Only under this condition can behaviorally derived quality spaces legitimately serve as evidence for the neuro-phenomenal mappings posited by NPS. Without such a structure-preserving causal link, behavioral similarity judgments could drift free from the neural patterns that supposedly mirror experience, severing NPS from its sole empirical access point and reducing its structural “mirroring” claim to an untestable coincidence.

Figure 3 summarizes the core structure of the argument. On the left, Premise 1 illustrates the neuro-phenomenal structural mapping assumed by NPS, linking neural patterns to phenomenal structure. At the top, Premise 2 (the measurability assumption) secures the possibility of inferring phenomenal structure through psychophysical methods. On the right, Premise 3

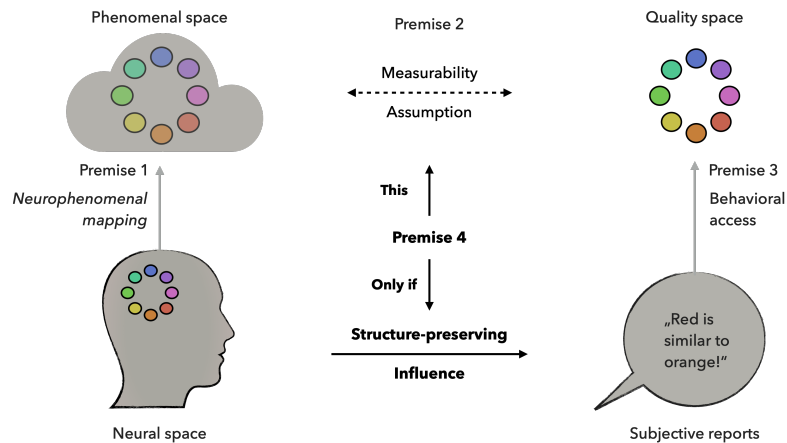


Figure 3: Illustration of our premises and conclusion

emphasizes that quality spaces are constructed from behavioral reports. At the center lies Premise 4, the crucial requirement: for these correspondences to hold together, the neural structures that mirror phenomenal structure must also shape behavior in a structure-preserving manner.

Premises 2 and 3 are widely accepted across consciousness science, and Premise 1 is the core assumption of NPS, which we grant for the sake of argument. The novel contribution is Premise 4. It shows that NPS is empirically tractable under Premises 1–3 *if and only if* the mirroring neural structures are linked to behavior in a structurally exploitable manner. Thus, neurophenomenal structuralists must focus on neural structures that influence behavior *in virtue of their structure*.

How, then, can we determine whether a neural structure genuinely shapes behavior in a structure-preserving way? This is a nontrivial challenge. Neural systems are extraordinarily complex, and structural similarities can arise by chance and lack functional significance. Putnam’s (1981) well-known example illustrates the danger: an ant’s meandering path in the sand may resemble the silhouette of Winston Churchill, but the match is purely coincidental—there is no causal link, no informational

connection, and certainly no representational use. This highlights a central problem for structural theories of representation: distinguishing *genuine* representational mappings from those that are trivial or accidental.

A widely endorsed solution centers on exploitability—a structure counts as representational only if a system can use it in a behaviorally relevant, structure-sensitive way (Artiga, 2023; Gładziejewski and Miłkowski, 2017; Piccinini, 2022; Ramsey, 2007; Shagrir, 2012; Shea, 2014, 2018, 2023). In neural terms, this requires that the relevant structural features be accessible to downstream processes and be *exploited* in generating specific behavior. Only then can a neural structure’s internal relational configuration do explanatory work, rather than merely coincide with the structure it purportedly mirrors.

This idea builds on Millikan’s consumer-based teleosemantics, which shifted the explanatory burden from the producer of a signal to its consumer. On this view, what makes a signal representational is not primarily how it was generated, but how it is used by systems that have evolved (or been trained) to interpret it in ways that support proper function (Millikan, 2004). In her account of intentional icons, structural isomorphisms—such as those between a bee dance and the location of nectar—gain semantic status only insofar as the consumers reliably exploit them, namely the bees that watch the dances to find nectar.

Our framework extends this insight to NPS. A neural structure only qualifies as a plausible NCCC if its internal organization does more than resemble phenomenal structure: it must be *structurally exploited* by downstream mechanisms that have evolved or learned to use it in generating behavior. Mere similarity is not enough. What matters is that the neural structure’s relational configuration plays a functional role in producing the similarity judgments and discriminations through which we measure phenomenal structure in the first place.

4 Which kinds of neural structures are promising NCCC candidates?

“Neural structures” are central to the NPS approach, but there are diverse types within the brain. These include anatomical structures, defined by the spatial relationships between cells; cause-and-effect patterns, which capture how neurons *can* influence one another; and activation patterns, as measured through functional brain imaging techniques, to name just a few. Which kinds of neural structures should NPS focus on? In this section, we present exploitation-based criteria to identify promising neural structures as candidate NCCCs for NPS.

4.1 Exploitable neural structure

The following two criteria can be understood as **preconditions for local neural structures to be exploitable**.

4.1.1 Sensitivity criterion

The Sensitivity Criterion simply asks whether changes in the relational configuration of a given neural structure can influence downstream processing in any meaningful way. It is essential to acknowledge that many neural properties—and the structures they form—do not significantly contribute to the brain’s information processing. Although neurons have a distinct color, the brain lacks mechanisms to track or use this feature for communication. If we were to paint neurons in rainbow colors, downstream processing would remain entirely unaffected. Such properties, even when measurable by us, cannot constitute an exploitable structure for the brain itself.

The first question to ask for a neurophenomenal structuralist, thus, is: are downstream systems *sensitive* to the relational configuration of a given neural structure? If the answer is no, the structure should be discarded as NCCC.

Anatomical organization offers a straightforward case study. The cortex exhibits intricate spatial layouts—such as the orientation maps in early

visual cortex, where neurons tuned to similar orientations are clustered together. Yet the spatial arrangement itself is not what downstream systems exploit. If a neuron were physically displaced while its connectivity and functional profile were preserved, neither the neuron nor its downstream targets would “notice”: its responses and all subsequent computations would remain unchanged. Functionally, it would still signal the same orientation.

Of course, physical displacement of cells would introduce many secondary effects, such as conduction delays. However, such effects could, in principle, be compensated for—for example, through changes in myelination. Once such secondary effects are controlled for, altering the spatial arrangement of neurons would leave downstream processing unchanged. This perspective aligns with the prevailing approaches to neural coding (Barack and Krakauer, 2021). Indeed, explanations of anatomical organization typically appeal not to computational function but to factors such as environmental statistics and developmental constraints (Lu et al., 2025) or wiring-cost minimization (Chklovskii and Koulakov, 2004).

In short, spatial layout—absent a mechanism that renders downstream systems sensitive to spatial relations—fails the Sensitivity Criterion and is therefore an implausible candidate for the neural structure that mirrors phenomenal structure under NPS.

Conversely, suppose the anatomical layout of a cortical map were to mirror phenomenal structure. In that case, perturbing that layout—while holding all functionally relevant relations constant—would still alter subjective experience correspondingly. Yet such perturbations would, by definition, leave subjective reports entirely unchanged. This could create profound dissociations between phenomenal structure and our only means of accessing it scientifically. Subjects would confidently report that their experiences remain unchanged, even when their underlying phenomenal structure would shift dramatically. This jeopardizes the epistemic foundation of consciousness science.

In summary, the Sensitivity Criterion serves as a necessary filter, excluding candidate neural structures whose internal relations, although measurable or even correlated with experience, lack direct causal and hence computational relevance for downstream systems and behavior.

4.1.2 Organization criterion

Building upon the Sensitivity Criterion—which establishes the requirement for a neural structure’s relations to exert some relevant causal influence on downstream processes—the Organization Criterion adds a further requirement: the *pattern* of relations within the structure must convey content systematically. That is, similar neural structures should map to similar reportable contents, while larger differences in structure should yield larger differences in content.

To elucidate this concept, we can draw upon Godfrey-Smith’s (2017) distinction between organized and non-organized representations. Non-organized (symbolic or nominal) systems lack any systematic alignment between the syntactic properties of the vehicle and the semantic properties of the content. Paul Revere’s lantern code provides a canonical example: “one if by land, two if by sea.” Here, the numerical difference between one and two lanterns has no inherent semantic significance; the mapping is purely arbitrary. By contrast, organized representational systems preserve similarity relations: the structure of the vehicle mirrors, at least to some degree, the structure of the content. If lantern *brightness* were used to signal the *size* of an approaching force, the system would be organized—minor changes in brightness would correspond to minor changes in estimated troop size. In such systems, vehicle-level similarity translates into content-level similarity.

Language illustrates the opposite pattern. Changing “Tomato” to “Tolato” destroys meaning entirely, whereas changing it to “Tornado” yields a radically different meaning—even though at the vehicle level, both modifications involve changes of comparable extent. Language-like representations are therefore non-organized: structural similarity between words does not predict similarity between contents.

Shea (2023) highlights the computational benefits of organized representations, explaining that they are more efficient, better suited for novel situations, and more resilient to errors. In an organized representational system, slight deviations in the signal do not significantly alter interpretation. For example, a slight change in the bee dance does not drastically

alter the interpreted distance to the nectar, ensuring the system is robust against noise.

NPS presupposes this kind of organization for the vehicles of conscious content: larger differences in neural structure should yield correspondingly larger differences in phenomenal structure, and therefore in the reportable contents through which we measure them. Without this systematic alignment, behavioral quality spaces would no longer track the relevant neural relations, and the core structural-mirroring claim of NPS would collapse.

Whether the brain in fact employs an organized representational format for conscious contents remains an open empirical question—one that can ultimately determine the viability of NPS itself.

The Organization Criterion challenges structural theories such as Integrated Information Theory (IIT), which posit that conscious experience is mirrored in the cause-effect structure (of the subset of the brain that has the highest Phi value—a measure of information integration), including not only active, but also inactive components and other dispositional properties (Tononi and Koch, 2015). This broad approach does not align with the Organization Criterion because it implies that changes to the brain’s causal structure may drastically change experiences without affecting downstream processes at all. Hence, there can be significant changes in phenomenal content that cannot be measured behaviorally, threatening Premises 2 and 3.

This can happen in several ways (Figure 4). First, adding simple local neural loops that change the causal structure of a system without altering downward processes. Second, consider inhibiting silent cells—those neurons that are inactive in the current state. This alters the brain’s cause-effect structure without influencing current processing (Bartlett, 2022), such that these inhibitions never translate into a difference in reported similarities either. Third, removing silent connections—those that do not currently participate in neural activity—also changes the cause-effect structure of the brain without impacting downstream processing. These examples demonstrate that not all parts of the cause-effect structure contribute to the downstream processes underlying quality spaces (see Doerig et al., 2019; Herzog et al., 2022; Tsuchiya et al., 2019 for further discussion on the dissociation

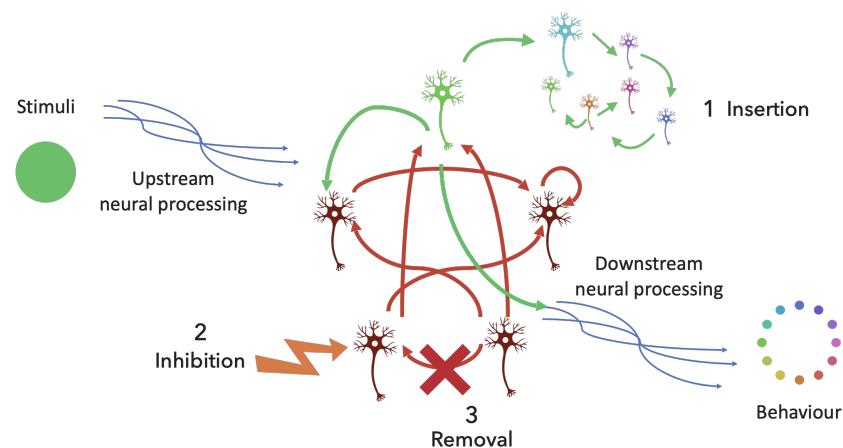


Figure 4: Illustration of dissociations between cause-effect space and quality space: 1.) Insertion of Irrelevant Loops: Adding neural loops that do not affect downstream processing but change the causal structure. 2.) Inhibition of Silent Cells: Inhibiting inactive neurons changes the causal structure without affecting ongoing processes. 3.) Removal of Silent Connections: Removing inactive connections changes the causal structure but not immediate behavior. Red arrows depict inactive connections, and green arrows depict active connections.

between causal structure and behavior).

In summary, the Organization Criterion requires that for a neural structure to be considered a valid NCCC candidate, changes in its structure must predictably and systematically influence phenomenal contents, and this systematic influence must be measurable through behavior under normal conditions. Slight changes in the structure should result in slight differences in content, while larger changes should lead to larger differences.

4.2 Are exploitable activation spaces really exploited

The previous two criteria focused on properties a candidate structure should have to be exploitable. These criteria challenge anatomical and cause-effect

structures as candidates for NPS. A more widely used option for neurophenomenal structuralists is to suggest that the structure of *neural activation patterns* mirrors phenomenal structures. Essentially, if two experiences are phenomenally similar, their corresponding activation should also be similar, while dissimilar experiences should produce more distinct patterns (Cichy et al., 2019; Kriegeskorte and Kievit, 2013). Such activation structures would satisfy our first two criteria and, in principle, be exploitable.

The following two criteria focus on whether an exploitable structure is actually exploited in the broader computational context of the brain.

4.3 Exploitation criterion

Shea (2014) puts the central issue neatly: “In many cases where there is a salient isomorphism, the correspondence is not in fact exploited.” In other words, it is not enough for a neural pattern to line up with some structure—the brain must use its relational pattern. The honeybee nectar dance illustrates this: the dance is richly organized, but bees never compare dances. Each dance is responded to individually, leaving its relational structure unexploited. A signal can thus satisfy the Organization Criterion yet fail the Exploitation Criterion if consumer mechanisms ignore its relational structure.

The same gap is evident in the brain. Neurons are often organized so that their firing varies systematically with stimulus features. An orientation-selective neuron in V1, for instance, responds more to its preferred angle than to orthogonal ones. In principle, this tuning curve carries similarity relations: a 20° tilt from the preferred orientation produces a stronger response than a 60° tilt. However, downstream systems do not exploit this relational gradient. Instead, they pool across many neurons or fire only if the incoming firing rate exceeds a threshold. Structural organization without structural exploitation, therefore, yields no genuine structural representation.

The transition from organization to exploitation is crucial. For example, a neural map already meets the Organization Criterion: similar activities in the map convey similar information. Place cells and grid cells in the

hippocampal–entorhinal system provide a clear example, forming spatial maps in which neighboring locations evoke overlapping activity patterns (Burgess et al., 1994; Moser et al., 2015; O’Keefe, 1976). At this stage, however, the rich relational information remains idle; only when these relations are explicitly exploited do we obtain a genuine structural representation.

Vectorial or relational coding provides the missing link between organization and exploitation. Boundary-, object-, and goal-vector cells are examples from the neuroscientific literature that exploit neural maps to compute relations rather than code single positions (Bicanski and Burgess, 2020; Høydal et al., 2019; Lever et al., 2009). Vector subtractions can encode distance between objects, while chains of such vectors can be used to simulate paths and discover shortcuts (Kurth-Nelson et al., 2023). Crucially, these relational codes are explicitly used to guide behavior in a structure-preserving way: in humans, hippocampal–entorhinal activity tracks distance to goals (Howard et al., 2014), and in bats, such goal-vector signals directly guide flight trajectories (Sarel et al., 2017).

The critical point here for our argument is that a brain region that merely mirrors a quality space is not sufficient as an NCCC. To qualify, downstream systems must actively exploit their structure in a structure-preserving way to guide behavior.

This becomes particularly clear when we revisit the winner-takes-all toy example introduced earlier (Figure 1). Here, brain region X encodes colors using overlapping groups of neurons, but a downstream mechanism selects the neuron with the highest firing rate and ignores the rest. In this case, the rich activation structure of region X is discarded and remains behaviorally irrelevant. The behaviorally derived quality space might still appear circular (e.g., downstream systems could infer that orange is more similar to yellow than to blue by tracking switching patterns between the “winning” neurons under changing lighting conditions or other temporal correlations). But critically, this quality space would not be causally grounded in the rich activation structure of region X.

The problem runs even deeper: according to the core hypothesis of NPS (Premise 1), perturbing the “losing” neurons—those not influencing behavior—would substantially change the subject’s experience if region X

mirrors phenomenal structure. As a result, the behaviorally derived quality space would no longer systematically track phenomenal structure, violating the measurability assumption.

In short, without assuming a structure-preserving influence, the argument for NPS falls apart, leaving it internally inconsistent and empirically untestable. In fact, such a ‘winner-takes-all’ strategy has recently been proposed in color processing (Zaidi and Conway, 2019), posing a real caveat for areas like V4—corresponding to color space—to serve as a structural representation and thus an NCCC for color experience.

In summary, the Exploitation Criterion requires that downstream processes actively use the relational information within activation patterns. Without such exploitation, dissociations become possible. A code can be richly organized, yet if downstream systems ignore its structure—say, through a winner-take-all readout or by only linking single place cells to respective rewards—this organization remains behaviorally inert.

4.3.1 Contextualization criterion

The Contextualization Criterion emphasizes the need to consider the broader computational context when determining how neural structures relate to content. As Shea (2018, p.96) points out: “A vehicle that carries correlational information about one state of affairs will usually carry information about many. Different downstream systems may be interested in different pieces of information: different correlations may be of use to each”. This remark is crucial because it clarifies that equivalent activation spaces can convey different meanings depending on how they are embedded in the brain’s overall processing.

A firefly’s light, for instance, signals “mating” to conspecifics but “prey” to predators. The vehicle itself is the same, but its meaning depends on the consumer. Similarly, in the brain, a motor signal that initiates movement when used by primary motor areas can simultaneously serve perceptual systems as information about expected sensory feedback. Again, what the signal conveys is fixed only in relation to the systems that exploit it.

Similarly, the Exploitation Criterion implicitly leads to the Contextualization Criterion. A neural pattern counts as a structural representation only if some downstream system uses its relational structure to guide behavior. However, “use” is impossible without interpretation by a consumer circuit shaped through learning or evolution. Lee’s (2019) IKEA-map example makes the point concrete: the same printed map can, in principle, match the layouts of many stores, yet it acquires a specific content—“this is the way to the desks”—the moment a shopper deploys it inside a particular shop. The shopper’s navigation goals fix which structural correspondences matter, by setting the map’s targets and accuracy conditions (see also Bickhard, 1999; Miłkowski, 2015). Change the context—use the map in a hardware store with the same floor plan—and the very same elements on the map may represent screwdrivers instead of sofas. In short, context also determines structural content; without that anchoring, structural similarity alone remains semantically underdetermined.

Accordingly, neural structures can also be used for entirely different things depending on their connections within the brain. Consider two structurally equivalent two-dimensional activation spaces, one for color processing and another for affect, as Figure 5 illustrates. Both spaces are structured by two opponent processing streams, resulting in two bipolar dimensions. For color, these dimensions are Red-Green and Blue-Yellow (De Valois and De Valois, 1993), while for affect, they are positive-negative valence and high-low arousal (Barrett and Russell, 1999). Despite both having a similar two-dimensional structure, their differing computational contexts—visual versus emotional processing—assign these spaces entirely different representational roles.

This is not to say that anything goes. Not all content can be mirrored in all structures. The structural ‘skeleton’ places constraints on which contents can be encoded. This point relates to Ashby’s (2022) notion of phenomenal schematics. Different kinds of experience come with different structural features that characterize them. A pressure experience always has both magnitude and direction—it has a vector-like structure. A temperature experience, by contrast, is scalar, varying only along a single hot-cold dimension. Because of this difference, a pressure experience cannot be re-

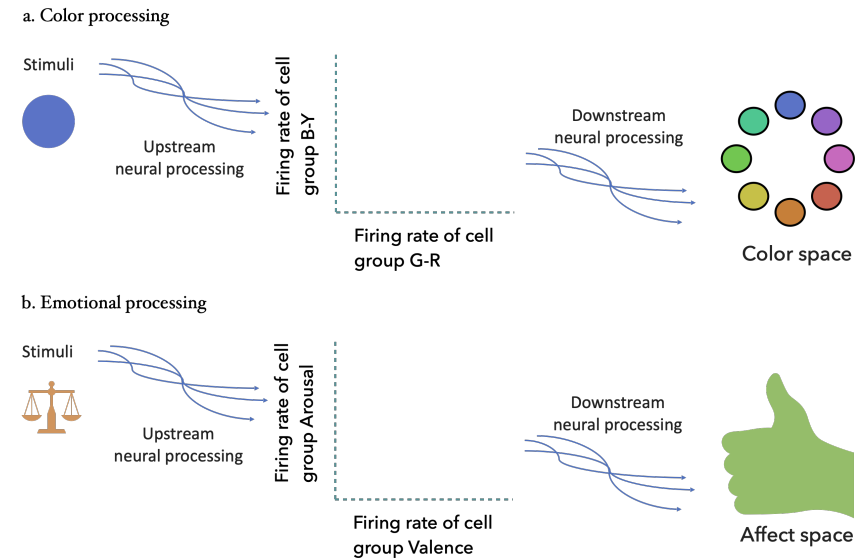


Figure 5: Illustration of the Ambiguity of Activation Spaces: The figure shows that two equivalent two-dimensional spaces can carry different contents depending on their context. In one case (a), the dimensions represent the opponent red-green and yellow-blue axes of the color space. In another case (b), the dimensions correspond to the opponent positive-negative valence and high-low arousal axes of affective processing. While both spaces share equal dimensionality, they are embedded within vastly different computational contexts—color perception versus emotional judgment.

duced to a temperature experience: the phenomenal schematics of pressure presuppose a second component that temperature lacks.

Neural activation spaces can be viewed similarly. Their structural skeleton—whether one-dimensional, two-dimensional, scalar, or vectorial—places some constraints on the kinds of contents they can support. A one-dimensional space cannot realize human color experiences, which require at least two opponent axes. However, beyond this schematic level, the structure by itself does not tell us what its dimensions mean.

The same two-dimensional activation space could, in principle, encode either color or affective states. Crucially, the brain cannot rely on the *experimental* context that happened to produce these patterns (for instance, that the subject was shown colored patches rather than emotional faces); only the experimenter has that information. For the system itself, the only way to fix content is through its broader computational context: which inputs feed into this space, which downstream circuits read it out, and which behaviors it helps control. If the activation space is tightly coupled to facial-expression processing and social-regulation circuits, this strongly favors an affective interpretation; if it is embedded in early visual pathways and guides color-dependent actions, a color interpretation is more plausible. On its own, the raw geometry of an activation space underdetermines its content.

In other words, phenomenal schematics narrow down the options by ruling out impossible mappings, but they cannot decide among the remaining possibilities. To determine content, we must ask: how is the information carried in this activation space actually used by downstream systems? Does the system treat the second dimension as “direction of pressure” or as “degree of arousal”? Only their functional integration can conclusively distinguish the two.

In summary, the Contextualization Criterion underscores the crucial role of computational context in determining the content of neural structures. Even if a local brain region’s structure aligns with a quality space and is exploited, understanding its role within the brain’s broader processing network is essential. Locally, without context, we cannot fully comprehend the representational role of an activation space, particularly its dimensions. This is why computational context is not merely an adjunct but a fundamental part of any structural NCCC.

5 Discussion

A key insight from our inquiry is that within the NPS framework, NCCCs must not only mirror phenomenal structures but also influence behavior

in ways that preserve their structures. To address this, we introduced four criteria to filter out neural structures that correspond only trivially to quality spaces and can not be appropriately exploited by downstream processes.

These criteria advance existing neurophenomenal mapping approaches by foregrounding the crucial role of the brain’s broader computational network. A neural structure qualifies as a viable NCCC only if downstream systems exploit it in ways that directly shape the similarity judgments from which the corresponding quality spaces are derived. Moreover, because this interpretive work depends on the structure’s position within the broader computational context, that context is not merely background—it becomes a constitutive part of the NCCC itself.

This analysis challenges local structuralist accounts. Local structuralists (e.g., Malach, 2021) hold that local lateral connections in sensory areas “implement” quality spaces for each sensory domain. This local conception is based on the idea that the activation patterns within these regions “exist in a similarity space, implemented in a neural activation space, the dimensions of which, as with size or color, give the experience its particular character” (Fleming & Shea, 2024). However, this raises the central issue highlighted by our Contextualization Criterion: what determines these dimensions in the first place? To understand what each dimension in a local activation space represents, one must examine how that space is embedded in the broader computational context that actually interprets and exploits it. By ignoring this content-constituting role of computational context, local structuralist theories risk proposing NCCCs that cannot account for the very contents they are meant to explain.

This concern connects directly to a fallacy identified by Brette (2018) in his critique of the “neural coding” metaphor. Coding models assume that neural activity locally encodes stimulus features. However, this inference depends on knowledge of the experimental conditions—information available to the experimenter, not to the brain. The brain must interpret its signals without access to their hidden causes. Local structuralists fall into a parallel trap: they infer what the dimensions of an activation space represent by appealing to the specific experimental manipulations that produced the activity patterns, thereby relying on information the neural

system itself does not possess. In doing so, they assign content prematurely and overlook the crucial role of computational context in fixing the representational role of an activation space.

6 Conclusion

NPS offers a promising approach to linking neural structures and conscious experiences, but its relationship to empirical testing needs further clarification. Our analysis shows that merely identifying correspondences between neural and phenomenal structures is insufficient. The relevant mapping must involve neural configurations that are not only structurally similar but also *exploitable* in ways that shape the behavioral similarity judgments from which quality spaces are constructed. To articulate this requirement, we introduced four exploitation-based criteria for identifying neural structures that genuinely contribute to the content of conscious experience.

These criteria reveal the critical role of computational context in determining the dimensions—and thus the content—of activation spaces proposed as NCCCs. They provide principled grounds for identifying and rejecting structural theories and NCCC candidates that underspecify how reported conscious content arises. Future work should refine these criteria and examine whether they extend to other non-structural neural coding schemes or whether defensible local accounts of content determination can be developed. Doing so will help clarify the relationship between neural activity and conscious experience and advance our understanding of NCCCs.

As a final remark, this paper emphasizes that addressing content-related issues is indispensable when evaluating theories of consciousness. NPS stands out by offering a clear hypothesis on how brain activity relates to phenomenal states—proposing that similar brain patterns give rise to similar experiences. This precision is often lacking in other theories but is crucial for their evaluation. We thus hope that all theories of consciousness one day provide an account of how their proposed NCCCs determine the structure of the experiences they are associated with.

Funding

Marlo Paßler was funded by the Deutsche Forschungsgesellschaft (DFG, German Research Foundation) - RTG 2386. Adrian Doerig was funded by ERC starting grant 101039524 - TIME.

References

- Artiga, M. (2023). Understanding structural representations. *The British Journal for the Philosophy of Science*, 728714. <https://doi.org/10.1086/728714>
- Ashby, B. (2022). Rainbow's end: The structure, character, and content of conscious experience. *Mind & Language*, 37(3), 395–413. <https://doi.org/10.1111/mila.12316>
- Baars, B. J. (1986). What is a theory of consciousness a theory of?—the search for criterial constraints on theory. *Imagination, Cognition and Personality*, 6(1), 3–23. <https://doi.org/10.2190/WJER-XABV-QM4W-KD6V>
- Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, 22(6), 359–371. <https://doi.org/10.1038/s41583-021-00448-6>
- Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, 8(1), 10–14. <https://doi.org/10.1111/1467-8721.00003>
- Bartlett, G. (2022). Does integrated information theory make testable predictions about the role of silent neurons in consciousness? *Neuroscience of Consciousness*, 2022(1), niac015. <https://doi.org/10.1093/nc/niac015>
- Bicanski, A., & Burgess, N. (2020). Neuronal vector coding in spatial cognition. *Nature Reviews Neuroscience*, 21(9), 453–470. <https://doi.org/10.1038/s41583-020-0336-9>
- Bickhard, M. H. (1999). Interaction and representation. *Theory & Psychology*, 9(4), 435–458. <https://doi.org/10.1177/0959354399094001>
- Brette, R. (2018). Is coding a relevant metaphor for the brain? *The Behavioral and Brain Sciences*, 42, e215. <https://doi.org/10.1017/S0140525X19000049>
- Burgess, N., Recce, M., & O'Keefe, J. (1994). A model of hippocampal function. *Neural Networks*, 7(6), 1065–1081. [https://doi.org/10.1016/S0893-6080\(05\)80159-5](https://doi.org/10.1016/S0893-6080(05)80159-5)
- Chklovskii, D. B., & Koulakov, A. A. (2004). Maps in the brain: What can we learn from them? *Annual Review of Neuroscience*, 27, 369–392. <https://doi.org/10.1146/annurev.neuro.27.070203.144226>
- Churchland, P. (2005). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology*, 18(5), 527–560. <https://doi.org/10.1080/09515080500264115>

- Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J. F., & Charest, I. (2019). The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. *NeuroImage*, *194*, 12–24. <https://doi.org/10.1016/j.neuroimage.2019.03.031>
- Clark, A., & Clark, A. (1993). *Sensory qualities*. Oxford University Press.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335. <https://doi.org/10.1016/j.tics.2016.03.006>
- Cohen, M. A., Nakayama, K., Konkle, T., Stantić, M., & Alvarez, G. A. (2015). Visual awareness is limited by the representational architecture of the visual system. *Journal of Cognitive Neuroscience*, *27*(11), 2240–2252. https://doi.org/10.1162/jocn_a_00855
- de Gardelle, V., Charles, L., & Kouider, S. (2011). Perceptual awareness and categorical representation of faces: Evidence from masked priming. *Consciousness and Cognition*, *20*(4), 1272–1281. <https://doi.org/10.1016/j.concog.2011.02.001>
- De Valois, R. L., & De Valois, K. K. (1993). A multi-stage color model. *Vision Research*, *33*(8), 1053–1065. [https://doi.org/10.1016/0042-6989\(93\)90240-w](https://doi.org/10.1016/0042-6989(93)90240-w)
- Doerig, A., Schurger, A., Hess, K., & Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition*, *72*, 49–59. <https://doi.org/10.1016/j.concog.2019.04.002>
- Feest, U. (2014). Phenomenal experiences, first-person methods, and the artificiality of experimental data. *Philosophy of Science*, *81*(5), 927–939. <https://doi.org/10.1086/677689>
- Fink, S. B., & Kob, L. (2023). Can structuralist theories be general theories of consciousness? In J. Hvorecký, T. Marvan, & M. Polák (Eds.), *Conscious and unconscious mentality* (pp. 112–129). Routledge.
- Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on neural correlates of consciousness. *Philosophy and the Mind Sciences*, *2*. <https://doi.org/10.33735/phemisci.2021.79>
- Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: Causally relevant and different from detectors. *Biology & Philosophy*, *32*(3), 337–355. <https://doi.org/10.1007/s10539-017-9562-6>
- Godfrey-Smith, P. (2017). Senders, receivers, and symbolic artifacts. *Biological Theory*, *12*(4), 275–286. <https://doi.org/10.1007/s13752-017-0276-4>
- Haberkamp, A., Schmidt, F., Hansmeier, J., & Glombiewski, J. A. (2019). Contamination-fear in subclinical obsessive-compulsive disorder: A further proof for no preferential processing of disorder-related stimuli. *Journal of Obsessive-Compulsive and Related Disorders*, *21*, 26–36. <https://doi.org/10.1016/j.jocrd.2018.11.004>
- Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: Long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, *24*(10), 826–837. <https://doi.org/10.1016/j.tics.2020.07.001>
- Herzog, M. H., Schurger, A., & Doerig, A. (2022). First-person experience cannot rescue causal structure theories from the unfolding argument. *Consciousness and Cognition*, *98*, 103261. <https://doi.org/10.1016/j.concog.2021.103261>
- Hesselmann, G., Darcy, N., Sterzer, P., & Knops, A. (2015). Exploring the boundary conditions of unconscious numerical priming effects with continuous flash suppression. *Consciousness and Cognition*, *31*, 60–72. <https://doi.org/10.1016/j.concog.2014.10.009>
- Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., Loftus, M. M., Staskute, L., & Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Current Biology*, *24*(12), 1331–1340. <https://doi.org/10.1016/j.cub.2014.05.001>
- Høydal, Ø. A., Skytøen, E. R., Andersson, S. O., Moser, M.-B., & Moser, E. I. (2019). Object-vector coding in the medial entorhinal cortex. *Nature*, *568*(7752), 400–404. <https://doi.org/10.1038/s41586-019-1077-7>
- Irvine, E. (2012). Old problems with new measures in the science of consciousness. *The British Journal for the Philosophy of Science*, *63*(3), 627–648. <https://doi.org/10.1093/bjps/axs019>
- Kammerer, F. (2021). The illusion of conscious experience. *Synthese*, *198*(1), 845–866. <https://doi.org/10.1007/s11229-018-02071-y>
- Kawakita, G., Zeleznikow-Johnston, A., Takeda, K., Tsuchiya, N., & Oizumi, M. (2025). Is my “red” your “red”? Evaluating structural correspondences between color similarity judgments using unsupervised alignment. *iScience*, *28*(3), 112029. <https://doi.org/10.1016/j.isci.2025.112029>
- Klein, C. (2010). Philosophical issues in neuroimaging. *Philosophy Compass*, *5*(2), 186–198. <https://doi.org/10.1111/j.1747-9991.2009.00275.x>
- Kleiner, J. (2024). Towards a structural turn in consciousness science. *Consciousness and Cognition*, *119*, 103653. <https://doi.org/10.1016/j.concog.2024.103653>
- Kob, L. (2023). Exploring the role of structuralist methodology in the neuroscience of consciousness: A defense and analysis. *Neuroscience of Consciousness*, *2023*(1), niad011. <https://doi.org/10.1093/nc/niad011>
- Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1882–1905. <https://doi.org/10.1037/0096-1523.25.6.1882>
- Kouider, S., Gardelle, V. d., & Dupoux, E. (2007). Partial awareness and the illusion of phenomenal consciousness. *Behavioral and Brain Sciences*, *30*(5–6), 510–511. <https://doi.org/10.1017/S0140525X07002919>

- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8), 401–412. <https://doi.org/10.1016/j.tics.2013.06.007>
- Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00245>
- Kurth-Nelson, Z., Behrens, T., Wayne, G., Miller, K., Luettgau, L., Dolan, R., Liu, Y., & Schwartenbeck, P. (2023). Replay and compositional computation. *Neuron*, 111(4), 454–469. <https://doi.org/10.1016/j.neuron.2022.12.028>
- Lau, H., Michel, M., LeDoux, J. E., & Fleming, S. M. (2022). The mnemonic basis of subjective experience. *Nature Reviews Psychology*, 1(8), 479–488. <https://doi.org/10.1038/s44159-022-00068-6>
- Lee, J. (2019). Structural representation and the two problems of content. *Mind & Language*, 34(5), 606–626. <https://doi.org/10.1111/mila.12224>
- Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, 29(31), 9771–9777. <https://doi.org/10.1523/JNEUROSCI.1319-09.2009>
- Lötsch, J., Kringel, D., & Hummel, T. (2019). Machine learning in human olfactory research. *Chemical Senses*, 44(1), 11–22. <https://doi.org/10.1093/chemse/bjy067>
- Lu, Z., Doerig, A., Bosch, V., Kraemer, B., Kaiser, D., Cichy, R. M., & Kietzmann, T. C. (2025). End-to-end topographic networks as models of cortical map formation and human visual behaviour. *Nature Human Behaviour*, 9(9), 1975–1991. <https://doi.org/10.1038/s41562-025-02220-7>
- Lyre, H. (2022). Neurophenomenal structuralism. a philosophical agenda for a structuralist neuroscience of consciousness. *Neuroscience of Consciousness*, 2022(1), niac012. <https://doi.org/10.1093/nc/niac012>
- Mead, A. (1992). Review of the development of multidimensional scaling methods. *Journal of the Royal Statistical Society Series D: The Statistician*, 41(1), 27–39. <https://doi.org/10.2307/2348634>
- Michel, M. (2024). The perceptual reality monitoring theory. In M. Herzog, A. Schurger, & A. Doerig (Eds.), *Scientific theories of consciousness: The grand tour* (1st).
- Millkowski, M. (2015). Satisfaction conditions in anticipatory mechanisms. *Biology & Philosophy*, 30(5), 709–728. <https://doi.org/10.1007/s10539-015-9481-3>
- Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy*, 86(6), 281–297. <https://doi.org/10.2307/2027123>
- Millikan, R. G. (2004). *Varieties of meaning: The 2002 Jean Nicod lectures*. MIT Press.
- Moser, M.-B., Rowland, D. C., & Moser, E. I. (2015). Place cells, grid cells, and memory. *Cold Spring Harbor Perspectives in Biology*, 7(2), a021808. <https://doi.org/10.1101/cshperspect.a021808>
- O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental Neurology*, 51(1), 78–109. [https://doi.org/10.1016/0014-4886\(76\)90055-8](https://doi.org/10.1016/0014-4886(76)90055-8)
- Overgaard, M. (2015). The challenge of measuring consciousness. In M. Overgaard (Ed.), *Behavioral methods in consciousness research* (pp. 7–20). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199688890.003.0002>
- Palmer, S. E. (1999). Color, consciousness, and the isomorphism constraint. *The Behavioral and Brain Sciences*, 22(6), 923–943, discussion 944–989. <https://doi.org/10.1017/s0140525x99002216>
- Paßler, M. (2023). The exclusionary approach to consciousness. *Neuroscience of Consciousness*, 2023(1), niad022. <https://doi.org/10.1093/nc/niad022>
- Pauen, M., & Haynes, J.-D. (2021). Measuring the mental. *Consciousness and Cognition*, 90, 103106. <https://doi.org/10.1016/j.concog.2021.103106>
- Piccinini, G. (2022). Situated neural representations: Solving the problems of content. *Frontiers in Neurobotics*, 16. <https://doi.org/10.3389/fnbot.2022.846979>
- Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10(2), 59–63. <https://doi.org/10.1016/j.tics.2005.12.004>
- Putnam, H. (1981). Brains in a vat. In *Reason, truth and history* (pp. 1–21). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625398.003>
- Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2022). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *The Behavioral and Brain Sciences*, 46, e261. <https://doi.org/10.1017/S0140525X22002849>
- Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511597954>
- 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*, 70(2), 581–607. <https://doi.org/10.1093/bjps/axx023>
- Rosenthal, D. (2010). How to think about mental qualities. *Philosophical Issues*, 20, 368–393.
- Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, 355(6321), 176–180. <https://doi.org/10.1126/science.aak9589>
- Shagrir, O. (2012). Structural representations and the brain. *The British Journal for the Philosophy of Science*, 63(3), 519–545. <https://doi.org/10.1093/bjps/axr038>
- Shea, N. (2014). VI—exploitable isomorphism and structural representation. *Proceedings of the Aristotelian Society*, 114(2), 123–144. <https://doi.org/10.1111/j.1467-9264.2014.00367.x>
- Shea, N. (2018). *Representation in cognitive science*. Oxford University Press.

- Shea, N. (2023). Organized representations forming a computationally useful processing structure. *Synthese*, 202(6), 175. <https://doi.org/10.1007/s11229-023-04373-2>
- Spener, M. (2022). Naive introspection in the philosophy of perception. *Review of Philosophy and Psychology*, 13(1), 29–45. <https://doi.org/10.1007/s13164-021-00597-8>
- Tee, J., & Taylor, D. P. (2020). Is information in the brain represented in continuous or discrete form? *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications*, 6(3), 199–209. <https://doi.org/10.1109/TMBMC.2020.3025249>
- Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668), 20140167. <https://doi.org/10.1098/rstb.2014.0167>
- Tsuchiya, N., Andriillon, T., & Haun, A. (2019). A reply to “the unfolding argument”: Beyond functionalism/behaviorism and towards a truer science of causal structural theories of consciousness. <https://doi.org/10.31234/osf.io/a2ms9>
- Tsuchiya, N., & Saigo, H. (2021). A relational approach to consciousness: Categories of level and contents of consciousness. *Neuroscience of Consciousness*, 2021(2), niab034. <https://doi.org/10.1093/nc/niab034>
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, 7(5), 207–213. [https://doi.org/10.1016/s1364-6613\(03\)00095-0](https://doi.org/10.1016/s1364-6613(03)00095-0)
- Vorberg, D., Mattler, U., Heinecke, A., Schmidt, T., & Schwarzbach, J. (2004). Invariant time course of priming with and without awareness. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation* (pp. 271–288). Psychology Press.
- Wilson, A. D., Tresilian, J. R., & Schlaghecken, F. (2010). Continuous priming effects on discrete response choices. *Brain and Cognition*, 74(2), 152–159. <https://doi.org/10.1016/j.bandc.2010.07.010>
- Yoshida, M. (1964). Studies in psychometric classification of odors (4). *Japanese Psychological Research*, 6(3), 115–124. <https://doi.org/10.4992/psycholres1954.6.115>
- Zaidi, Q., & Conway, B. (2019). Steps towards neural decoding of colors. *Current Opinion in Behavioral Sciences*, 30, 169–177. <https://doi.org/10.1016/j.cobeha.2019.10.011>