

A structural constraint on neural correlates of consciousness

Sascha Benjamin Fink^{a,b,c}  (sfink@ovgu.de)

Lukas Kob^{c,d,e}  (lukas.kob@ovgu.de)

Holger Lyre^{a,b,c}  (lyre@ovgu.de)

Abstract

Researchers on the neural correlates of consciousness (NCC) need to distinguish mere statistical NCCs from NCCs proper. Some neural events may be co-occurrent, probabilistically coupled, or coincidental with a type of conscious experience but lack any deeper connection to it, while in other cases, the relation between neural states and a type of experience hints at a strong metaphysical relation, which distinguishes such NCCs proper from mere statistical NCCs. In order to address this issue of how to distinguish NCCs proper from mere statistical NCCs, we propose a position we call *neurophenomenal structuralism*. The position hinges on the uncontroversial idea that phenomenal experiences relate to each other in degrees of similarity and difference. These complex structures are used to identify and individuate experiences in the methods of neuroscience, psychophysics, and phenomenology. Such individuation by structure leads to *phenomenal holism*, which has implications for how to investigate consciousness neuroscientifically and generates a constraint by which we can distinguish NCCs proper from mere statistical NCCs: the *structural similarity constraint*. Neural activation must preserve the structure governing the domain of experiences it is associated with in order to count as that domain's NCC proper. Any activation that fails to preserve phenomenal structure fails to be an NCC proper. We illustrate how this constraint works with a study by Brouwer & Heeger (2009) as an example.

Keywords

Color · Holism · Individuation · Neural correlates of consciousness · Neurophenomenal structuralism · Representational geometry · Structuralism

This article is part of a special issue on “The Neural Correlates of Consciousness,” edited by Sascha Benjamin Fink and Ying-Tung Lin.

^aPhilosophy Department, University of Magdeburg

^bCenter for Behavioral Brain Sciences (CBBS), University of Magdeburg

^cHumboldt-Universität zu Berlin, Research Training Group 2386 “Extrospection”

^dBernstein Center for Computational Neuroscience, Charité – Universitätsmedizin, Berlin

^eHumboldt-Universität zu Berlin, Faculty of Philosophy, Berlin School of Mind and Brain

1 Introduction: Statistical vs. proper correlates of consciousness

One of the primary goals in the neuroscience of consciousness is to find the neural correlates of consciousness (NCCs). Generally speaking, correlation is a measure of the dependence between the values of variables. In NCC research, the variables at issue are neural versus phenomenal goings-on.

At its heart, correlation is a statistical notion. It can be positive, negative, and comes in degrees—all the way to where it nearly touches zero. In most samples, one will find variables that correlate to some non-zero degree. In finite and especially in small samples, the correlation may even be statistically strong despite a lack of meaningful connection between the correlating variables. Pearson (1897) called these “spurious correlations.”

Spurious correlations raise a general worry: If statistical correlations between phenomenal and neural goings-on can be found without any meaningful connection between them, how valuable are the contributions of NCC research to finding the neural substrate of consciousness? Statistical procedures may help to minimize spurious correlations, but they cannot rule them out completely. Beyond spurious correlations, there are other problems too. For instance, some statistical correlates are reliably correlated with a specific conscious experience c such that they can be used as indicators of c but are nevertheless not c 's foothold in the observable physical world. This may be because they occur before or after c (see also Aru et al., 2012), or are simultaneously co-caused by whatever causes c , or are just a common but not universal or necessary side-effect of c . So there might be a broad range of neural states that are reliably correlated with c , but do not qualify for being a neural substrate of c . Their correlation with c —however strong—is not indicative of any strong metaphysical connection (supervenience, grounding, identity, etc.). The correlation is merely statistical. While these may be useful to identify, we will consider only more-than-merely-statistical correlates as the *NCC proper*.

If NCC research wants to contribute to finding the neural substrate of consciousness, additional constraints beyond correlation become attractive. These allow us to reject certain recorded statistical correlates as candidates for the NCC proper.¹ In our view, most of the currently discussed constraints are limited in their applications or have other drawbacks.² So here, we want to motivate a constraint that stems from the methodology of NCC research itself, inspired by an already prominent emphasis in this field on the structures of phenomenality.

¹The main options currently under discussion as additional constraints are: the capacity to account for phenomenal features (Seth, 2009; Seth & Edelman, 2009), synchronous occurrence with the phenomenal experience (Aru et al., 2012), being systematically entailed by a theory (Hohwy & Seth, 2020), being necessary and sufficient (Crick, 1995), and being minimally sufficient (Chalmers, 2000).

²See, e.g., Chalmers (2000) for a critique of Crick (1995) and Fink (2016) for a critique of Chalmers (2000).

The proximity of NCC research to psychophysics³ underpins a position we call *neurophenomenal structuralism*, which yields a *structural similarity constraint* on NCCs. This constraint eliminates a broad range of statistical correlates as candidates for NCCs proper.

In section 2, we offer an overview of the emphasis on structures of experience in phenomenology and psychophysics, and show how they can be useful in individuating and identifying experiences. This motivates the position of *neurophenomenal structuralism*, which we sketch out in section 3. This position rests on the idea that (i) each experience is, in scientific contexts, structurally individuated and that (ii) there is a systematic relation between phenomenal structures and structures in the neural domain. A strong metaphysical reading of these twin ideas is possible: Every phenomenal character is exhaustively individuated by its relations and not, as qualia theorists hold, by any intrinsic phenomenal properties. We are sympathetic to this reading, but focus here on a methodological point: In section 4, we propose that a structural morphism of a neural substrate with phenomenal spaces is an additional constraint on NCCs proper. If phenomenal character is individuated structurally, and if phenomenal structure is mirrored in neural structure, then NCCs proper must preserve structural relations of similarity and dissimilarity between experiences, while mere statistical correlates likely do not. We call this the *structural similarity constraint* (SSC). In subsection 4.3, we illustrate this constraint in action with an analysis of a study by Brouwer & Heeger (2009) on color perception. However, we believe our point extends to all phenomenal domains.

2 Structures in phenomenology and psychophysics

Phenomenologically, experiences (e.g., as of colors, odors, tones) relate to each other in terms of similarity and dissimilarity. For example, seeing the color red resembles seeing orange more than it does seeing green; the smell of parmesan resembles the smell of vomit more than it does the smell of mint; the sound of the chord *C-maj*⁷ resembles *A-min*⁹ more than it does *F \sharp -maj*⁷.

For a given domain of experiences (like a sensory modality or parts of it, e.g., colors in vision), we can map out these relative similarities and dissimilarities between types of experience across the domain. If we represent the resulting order geometrically, relative dissimilarity is associated with distance (see the left hand side of Fig. 1 in section 4.3 for an example). That is why green is located opposite red on most hue circles: Independently of the concrete visual scene, whenever one experiences red, it is the hue most dissimilar to actual or possible green experiences.⁴ Some depictions of such orderings are widely known: The Munsell

³To some degree, NCC research can be seen as a descendent of what Fechner (1860) called “inner psychophysics.”

⁴Billock et al. (2001) apparently produced reddish-green experiences in experimental settings (see also Hardin, 2014). This means that red is not *always* most distant to green. But this does not speak against our overall point: We are not arguing for a specific ordering of experiences; we are

color spindle and Newton's or Goethe's color wheel are still widely used in art classes. Many such attempts to map out a phenomenal domain predate our current scientific approaches in psychology, psychophysics, and neuroscience, especially when it comes to color (Kuehni, 2003). Within the sciences, psychophysics is the field most prominently concerned with charting the structures that govern the relative similarities between experiences, by investigating "just noticeable differences" (JNDs) between them. By developing a wide range of extrospective methods, psychophysicists became able to represent these ordering structures in a finer-grained manner than artists had done introspectively (Gescheider, 1997; Lawless, 2013).

Granted, experiences differ—but what makes an experience specifically *this* one? This is a question of individuation. Individuation is differentiation at the limit. By differentiating *a* from *b*, we say why these two entities are not identical; by individuating *a*, we say what differentiates *a* from all other entities. So if we know the individuation criteria for some *a*, we can say why *a* is not identical with any other entity. Individuation is thus a pathway to identification: If some *x* fulfills all individuation criteria for *a*, then $x = a$. How reliable this pathway is depends on the extent to which we can *know* whether *x* fulfills all individuation criteria. Some metaphysicians contemplate individuation criteria that are brute, irreducible, or even inaccessible. Here, we will bracket such metaphysical debates and focus on a methodological question instead: How can *we* individuate? Given an empirical method *m*, we can ask: How does *m* help us to individuate those entities accessible via *m*? For psychophysics specifically, we can ask: How do we individuate experiences psychophysically?

Differentiation of experiences is a key element in psychophysics: Psychophysicists measure our ability to match or discriminate experiences in experimental settings in order to extract the relations by which humans order their experiences based on relative similarity judgments implicit in their sorting behavior. For example, *a* is just like *b*, but *a* is just noticeably different than *c*, while *b* is not, etc. Such triangulating judgments allow us to model how a phenomenal domain is ordered, at least relative to the psychophysical methods used. We can call these orderings *phenomenal structures*⁵ if they are meant to reflect relations between types of phenomenal characters.⁶ In experiments, psychophysicists will relate these phenomenal structures to non-phenomenal goings-on. Most often, the experiences to be matched or discriminated are elicited by external stimuli. The phenomenal struc-

simply claiming *that* they are ordered. Finding out the specific ordering that governs a domain might be a highly complex task. It might even turn out that a 3+-dimensional color ordering, as Niederee (2010) suggests, is more adequate than a 3-dimensional one in explaining, e.g., the effects produced by Billock et al. (2001) or color relations in tetrachromacy.

⁵These differ from the structures *in* an individual phenomenal experience token, such as the distribution of colors and shapes in a single visual impression.

⁶That is, they are not meant to reflect the relations between the causes of these experiences outside of the organism or between the stimuli outside of how they appear to us. Obviously, these do matter for eliciting experiences, but psychophysical spaces are not always interpreted as merely reflecting an ordering in the external world.

tures are thereby automatically tracking structures of stimuli as well.⁷ However, we may also investigate experiences that do not correspond to any external or specific stimulus. Chimerical colors (Churchland, 2005) are a case in point, i.e., color sensations of afterimages that do not correspond to any reflective stimuli but that can nevertheless be systematically induced in a variety of ways.

Understood geometrically, these phenomenal structures comprise *phenomenal spaces*. These capture the difference and similarity structures of all phenomenal characters in a domain. Each methodologically accessible phenomenal character has a location in a phenomenal space. Such spaces are spanned by certain abstract dimensions. Each individual in this space has a location which corresponds to coordinates along these dimensions. The labels we give these dimensions reflect the features in respect to which individuals differ or resemble each other. For colors, such orderings are called *color spaces*⁸ (Kuehni, 2003; Kuehni & Schwarz, 2008) and hue, saturation, and lightness are dimensions along which colors can differ.⁹

In some cases (e.g., hue, saturation, and lightness) the dimensions of variability are introspectively available. However, in psychophysics, the dimensions are usually associated with features that are only extrospectively measurable. In “outer” psychophysics, the dimensions of the phenomenal spaces are related to properties of the stimuli, e.g., wavelengths, amplitudes, intensity. In “inner” psychophysics, these dimensions are associated with internal features of the system, i.e., neural features.¹⁰ Accordingly, some color spaces are supposedly grounded in neural activation itself, for example the color-opponent processes in ganglion cells or the LGN (on DKL color space, see Derrington et al., 1984; Kuehni & Schwarz, 2008, p. 148), or are linked to the cone responses of the retina (Kuehni, 2003, pp. 164–168).

It is reasonable to propose (i) that human experience possesses some kind of order (even though this order may be highly complex or otherwise hard to model ad-

⁷Philosophers such as Decock (2006) therefore suggest that these structures are redundant and must be interpreted as purely physical.

⁸The terms “color space” and “color solid” are both used. In Churchland (2005) and Kuehni & Schwarz (2008), the relation between color solid and space appears to be that psychophysics provides us with highly complex and asymmetrical shapes, the color solids, which represent the order in which all perceivable colors relate to each other. These solids are a subspace inside a larger color space. However, one can think of phenomenal spaces themselves as determined by the structure governing phenomenal characters. In this understanding, the solid need not inhabit a space independent of it, but is simply a finite space.

⁹Such color spaces derived from matching experiments have proven useful for industrial purposes (e.g., CIE color space; Kuehni & Schwarz (2008) 141–142). For example, when it comes to choosing colors for buttons that initiate opposite actions (like start/stop), it is useful to choose colors from opposing extrema of a color space. Color spaces have been operationalized in many different ways, and numerous color space models are therefore available, differing in their experimental origin, their metrics, and the explanations they provide. Still, in each of these spaces, colors can be individuated structurally, namely by their location inside the overall space.

¹⁰This distinction was first drawn by Fechner (1860).

equately), and (ii) that psychophysics methodically approximates this ordering.¹¹ At the limit of psychophysical inquiry, an individual type of experience (like Red₄₂) is individuated by its relative difference to all other experiences in a domain, i.e., by its location in the overall structure of a phenomenal space, as established by some psychophysical method. This principle will hold not only for color, but for all modalities.¹²

3 Neurophenomenal structuralism

In this section we develop our proposal of *neurophenomenal structuralism*. It rests on two assumptions. First, it is the location in a structure, not an intrinsic property, that individuates a phenomenal character. This view could be called *phenomenal structuralism* and is introduced in subsection 3.1. Second, phenomenal structure is mirrored in neural structure. This view leads to *neurophenomenal structuralism* and is explored in 3.2. A remarkable consequence of phenomenal structuralism (and therefore also of neurophenomenal structuralism) is *phenomenal holism*, which we develop in 3.3.

3.1 Individuation by location in a structure

The core idea of neurophenomenal structuralism is that experiences are structurally individuated. In contrast to this, some philosophers hold that experiences are intrinsically individuated. According to them, what makes an experience have a particular phenomenal character is not reliant on anything else; it is a non-relational, non-extrinsic phenomenal property. This is the essence of the idea of qualia as introduced by C.I. Lewis (1990). According to such intrinsicists, structures—relations to other experiences, to stimuli, to our cognitive capacities, to our conceptual schema—might help us identify what experiences we are having, but they do not give the experiences their phenomenal character (Levine, 1995). A broad range of philosophers (us included) disagree with this metaphysical position for several reasons. But here, we want to focus only on the methodological side. Suffice it to say that, in the intrinsicists' picture, one must somehow access these intrinsic but individuating properties in order to truly make out the kind of experience one is dealing with. However this is supposed to work, such a kind of access can hardly be made plausible in a scientific context. Ultimately, science deals with

¹¹There are complications, however. For example: How do different color spaces relate to each other and to human experience? Which color space reflects the true order of human color experience? And so on. These are valuable questions, but we lack the space to expand on them here. Instead, we focus on methodology and are happy to accept that color spaces exist relative to methods for assessing phenomenal orderings.

¹²See also Clark (2000) and Rosenthal (2015). For a comparable take on olfaction, see Young et al. (2014), and for the auditory system, see Isaac (2018). Color is simply the most studied example by far.

structures. In the science of psychophysics, experiences are individuated by their structural location, as stated in the previous section.

But what is a structure, specifically? A structure is a set of relata with a set of relations imposed on them.¹³ A structure in this sense is a tuple $\Sigma = \langle a, R(a) \rangle$ consisting of a carrier set or domain $a = \{a_1, a_2, \dots, a_k\}$ of k elements or entities a_i equipped with a set of n -ary relations $R(a)$. Therefore, to capture a domain structurally is to individuate the entities or relata only via the relations in which they stand.

Now, given two domains, we can map entities in one domain onto entities in the other. Some mappings will also preserve the structures in a domain, and such structure-preserving mappings are known as *homomorphisms*.¹⁴ Of particular interest are bijective homomorphisms, called *isomorphisms*. If two structures are isomorphic, then they are structurally indistinguishable. In other words, any structure is defined only up to isomorphy. This matters if we care about individuation, i.e., if we care about the question of what determines that something is specifically *this* and not something else. Do we need more than interrelations to make out what something is, or that it is *this*?

Structuralists say that structure is enough. All we can say for a given entity is that it is in *this* position in the structure. Two views contrary to this have been prominent in traditional metaphysics: According to intrinsicists, entities are individuated by intrinsic properties (properties that an object may have irrespective of the existence of other objects); according to haecceitists, identity is grounded in *haecceity* (a sort of “identity property” or “primitive this-ness”). Structuralists deny both views: the relata in a relational structure possess neither any intrinsicity nor any haecceistic individuality, since they are captured only by their location in the overall structure. There is nothing to specifying what something is over and above stating its location in a structure.

This basic structuralist claim accords with the first assumption of phenomenal structuralism: In the domain of the phenomenal, insofar as it is captured by psychophysical methods, a phenomenal character is identified by its location in a structure.¹⁵ In and of itself, this assumption motivates a *phenomenal struc-*

¹³Strictly speaking, this is a *relational structure*. Structured sets may come equipped with either functions or relations or both. If a structure contains no relations, it is an algebraic structure; if it does not contain any functions, it is a relational structure (or system of relations). Throughout this paper we will use the term *structure* in the sense of relational structure.

¹⁴Two relational structures $\Sigma^a = \langle a, R(a) \rangle$ and $\Sigma^b = \langle b, S(b) \rangle$ are of the same type, if for every relation $R_j \in R$ there is a corresponding relation $S_j \in S$ with the same arity. Given two such structures, then a mapping $h: a \rightarrow b$ is a homomorphic mapping or *homomorphism* from Σ^a to Σ^b , if for every relation R_j and any elements $a_i \in a$ the following implication holds: $R_j(a_1, a_2, \dots, a_k) \Rightarrow S_j(h(a_1), h(a_2), \dots, h(a_k))$.

¹⁵Wiese (2018) and Tsuchiya et al. (2021), among others, have made remarks which partly mirror our approach. However, we do not straightforwardly share most of the additional constraints made by these authors. For example, we currently remain neutral on whether the structure must be considered metrical (as Wiese does) and remain neutral on whether this structure must be captured in category theory (as Tsuchiya et al. suggest).

turalism which can be accepted even by anti-physicalists because it does not yet say anything about phenomenality's relation to any physical stimulus or neural matter. In NCC research, however, we have to consider the neural domain alongside the phenomenal. The connection between the two is the subject of our next section.

3.2 Phenomenal structures are mirrored in neural structures

Scientific methodology captures the structure of a domain. Psychophysics attempts to do so for phenomenal characters. But it does more than this: It also relates the structure of experiences either to the structures of external stimuli (in outer psychophysics) or to the structures of neural goings-on (in inner psychophysics). NCC research relates most closely to inner psychophysics because it aims to know which phenomenal characters relate to which neural activations. Therefore, we shall consider two domains and their interrelationships:¹⁶

- The *P*-domain of phenomenal states and processes (i.e., qualitative experience)
- The *N*-domain of neural states and processes (i.e., neural activity)

Because processes can be considered transitions between states, the various domains are sufficiently modeled by the respective state spaces and the dynamics defined on them:

- The *P*-space (i.e., phenomenal space or quality space)
- The *N*-space (typically a neural activation space)

Finally, since we are adopting a structuralist perspective, we shall focus on the respective structures of the two spaces:

- The *P*-structure (governing *P*-space)
- The *N*-structure (governing *N*-space)

¹⁶In experimental contexts, experiences are often but not always elicited by specific sensory stimuli. We might therefore consider a third domain of sensory stimuli. There are, however, exceptions. Afterimages or metamers can be elicited by various stimuli. Additionally, some experiences are elicited solely by interventions on the brain and are not tied to a specific *sensory* stimulus. These include experiences elicited by pharmacological interventions in psychedelic research, by direct brain stimulation, or by transcranial magnetic stimulation. The connection of experiences to stimuli is thus likely neither exhaustive nor exclusive, but certainly complex. Here, we focus on the search for NCCs in the context of inner psychophysics and therefore ignore the connection to external stimuli for the moment.

Now, the second assumption of neurophenomenal structuralism says that P -structure is mirrored in N -structure. This needs to be explained in more detail. The notion of mirroring requires us to consider the interrelationships between P - and N -structures in terms of their *structural similarities*. In general, two structures A and B are structurally similar if the corresponding relations in A and B have the same number of arguments. Paradigmatic cases of structural similarity are between maps, pictures, or sculptures and what they represent.

Of course, such structural similarity is a second- rather than a first-order similarity. As several authors have pointed out (O'Brien & Opie, 2004; Shepard & Chipman, 1970), first-order similarity consists of shared properties in both the source and the target domain. For example, colored chips are used by interior designers to select the intended color for painting a room. In the case of second-order similarity, however, it is only the relations between the relata (i.e., the structural properties) that are shared (or "mirrored"). In a weather map, for instance, the spacing of isobars corresponds to pressure gradients in the atmosphere. In an analogous manner, second-order similarity holds for the structural similarity between the phenomenal and the neural. In the case of color, to say that P -structure is mirrored in N -structure does not, of course, mean that neurons or neural activations are in any (bizarre!) sense colored, but rather that changes in neighbouring color sensations (as encoded into color space) correspond to similar changes in neural activity (as encoded, for example, in the neural activation space).

The strongest form of structural similarity between P and N would be an isomorphism. However, we must account for neural plasticity and so cannot ignore the multiple realizability of phenomenal states by neural states. Multiple realizability is also at the heart of the NCC definition as introduced by Chalmers (2000) in contrast to Crick (1995): Any specific neural activation that is an NCC counts only as sufficient, but not as necessary, for its associated phenomenal experience. Therefore, rather than isomorphism, we should expect a *surjective homomorphism* from N to P , since a surjective mapping is a many-to-one mapping.¹⁷ Surjective homomorphisms are also known as *epimorphisms*.

The requirement of a surjective homomorphism accounts for the idea that phenomenal structure must be preserved when mapped onto the neural domain. The converse, however, need of course not be true. Plenty of structures can certainly be found in the neural domain that do not feature in phenomenology (it is, for instance, of no importance what the structure of the neurons' physical masses is). Still, in order to provide a neural account of consciousness, as NCC researchers aim to do, all phenomenal structures must find a neural correspondence, but the

¹⁷Given two sets A and B , an injective mapping from A to B is a one-to-one mapping where we may also have elements of B without matching A -elements. A surjective mapping from A to B means that every B -element has at least one matching element of A , but maybe more than one. Hence, in general, a surjective mapping is a many-to-one mapping. A mapping that is both injective and surjective is called bijective. Therefore, a bijective mapping amounts to a one-to-one correspondence where each element of one set is paired with exactly one element of the other set.

$P \rightarrow N$ correspondence can be one-to-many.¹⁸ Conversely, a surjective, many-to-one $N \rightarrow P$ mapping accounts for the notorious multiple realizability of the phenomenal by the neural, as neural connections may be functionally replaced by others due to neural plasticity. This means, for instance, that even after certain neurons have died, one may continue to experience colors the way one used to.

3.3 Phenomenal holism

As we noted above, the phenomenal character of an experience is traditionally considered to be an intrinsic property of an experiencing subject's mental state (i.e., a *qualia*). Qualia enthusiasts like Goff (2017) claim that we identify a phenomenal experience by grasping this intrinsic property via introspection. Neurophenomenal structuralism, by contrast, claims that to individuate a phenomenal character means to capture a relational rather than an intrinsic property.¹⁹

To illustrate this, consider color phenomenology. We should think of a color space not as representing *absolute* values, but rather as a means to represent *relations between colors*.²⁰ Similar views have been expressed by several other authors. Isaac (2014) argues that we use relations between colors to flexibly represent changes and relations occurring in the environment. Papineau (2015) reminds us that even though humans are said to be capable of distinguishing millions of colors, we are not able to distinguish them in an absolute fashion. While subjects are typically not able to distinguish between, say, Blue-27 and Blue-28 if shown one after the other, they might very well be able to discriminate them if shown side by side. Gert (2017, p. 6) speaks of "brutely relational processes [that] yield relational information directly, without *inferring* it from non-relational information about the relata." Finally, Clark (2000, p. 13f) says:

When one sees a patch of orange, the experience seems to involve an intrinsic monadic quale: the quale orange. But this experience is an illusion. The facts in virtue of which that experience is an experience of orange, and not of some other quality, are all relational facts. [...] Qualitative content may be relational even though it is not invariably experienced as relational. [...] the facts in virtue of which a sensation has the particular qualitative character it has are all relational facts.

¹⁸This opens an attractive door for reductionism, but we will leave this out for lack of space. Additionally, there may also be a non-reductive reading of neurophenomenal structuralism.

¹⁹And because structure on P -space has a correspondence in N -space, we might also do so retrospectively.

²⁰Note, however, that in the color realism debate the term "relationalism" is frequently used to indicate that color is a property emerging in the relation between a physical object and a perceiver (e.g., Chirimuuta, 2015, ch. 3). This is not what we have in mind here (although we are also not explicitly opposed to this view). Our position is rather that color experiences are relationally individuated among themselves, even if the colors are hallucinated or dreamt of.

The “relational facts” that Clark has in mind here might pertain to the external regime of color stimuli as well as to goings-on in the neural substrate (e.g., color opponency).²¹ While this is likely true, we here focus only on the relations governing the color space itself and for the moment ignore its relation to external stimuli²² or a neural substrate: The phenomenal character of each color is what it is due solely to its relation to all other colors. This is what we have called phenomenal structuralism.

If a type of experience in a domain is individuated by its relations of graded similarity and difference to *all* other types of experience in that domain, as structuralists claim, then this has the remarkable implication of *phenomenal holism*: The *P*-structure *as a whole* determines each internal relatum. Hence, phenomenal holism is implied by phenomenal structuralism. Consider again color phenomenology: Any single color experience is constituted via its location within the whole color space. As Clark (2000, p. 18) puts it:

So orange can at best be defined as something like a reddish yellow, equally similar to red as to yellow, complementary to turquoise, more similar to red or to yellow than it is to turquoise, and so on.

To draw an analogy: Phenomenal structuralism implies phenomenal holism in much the same way as functional role semantics leads to meaning holism (Block, 1998). Meaning holism is concerned with the meaning of words in a language or the intentional contents of beliefs and thoughts. Each word has a specific meaning only when we consider the meaning of all other words in the language because they stand in inferential and semantic relations to each other. In the linguistic context, holism is widely discussed; much less so when it comes to phenomenality. And this is precisely our project.

Meaning holism has been criticized, but these criticisms do not transfer to phenomenal holism. One standard argument against meaning holism is that meanings may become unstable, may vary among subjects, and thus cannot reliably be shared (Fodor & Lepore, 1992). This argument is by no means uncontroversial and can be countered (Block, 1998). But in any case, this type of argument has no force in relation to holism about phenomenality. On the contrary: We should expect considerable variability in the individual *P*-structure for members of a given species (Fink, 2018; Hohwy, 2011). From the perspective of neurophenomenal structuralism, this is not surprising and does not pose a problem: The structure governing an individual’s experiences is reliant on the structures of that individual’s nervous

²¹Clark (2000, p. 138) argues that “[...] in sensory systems the root relation from which order derives is not similarity, but difference. Instead of qualitative identity as the primitive term, the neural reality would be better reflected by using discriminability: the detection of differences,” where these differences are external.

²²Of course, we understand that color perception are ultimately anchored in external stimuli. But the relationship between stimuli and phenomenal character is tangled and complex. Here, the focus is on NCC research, so we will bracket this complication.

system; but neural structures vary inter-individually because the development of biological neural systems depends on individual and environmental circumstances. Only by knowing a subject's full *P*-structure can we know what specifically they are indeed experiencing now. All else is approximation.

In the case of color, such variations are known. People vary not only in the make-up of their retinæ, but also in the size and interconnections of their visual cortices. Neurophenomenal structuralists must therefore predict varying color experiences. And these show themselves in how people diverge in color judgments in everyday scenarios (Hofer et al., 2005). Additionally, high inter-observer variability in placing the unique hues has been empirically observed (Kuehni, 2004). Or consider color blindness, for instance the case of red-dichromacy (protanopia): If an entire way of eliciting color experiences is lost, this must show itself in the individual colors experienced.²³ Or how about adding new cones to the retina? Experimental data indicate that there are human tetrachromats who pick a higher number of unique hues than their trichromatic conspecifics (Jameson et al., 2001).

Phenomenal holism adds another twist. Subjects with red–green color blindness do not only fail to distinguish red and green. If each color is individuated by its relation to all others, they will experience *all* colors differently, since their color space structure differs significantly from the color space structure of trichromats.

Neurophenomenal structuralism is in stark contrast to the classical qualia view. There, we identify experiences by accessing intrinsic properties which are available only via introspection (Goff, 2017). Phenomenal character can thus be assessed atomistically. All you ever experience could be one single color, but you would still know which color it was. Structuralism, instead, leads to phenomenal holism: Each experience has its phenomenal character only in virtue of other experiences having specific characters. Only by accessing these relations can we identify phenomenal characters. Therefore, even if (metaphysically) all one could ever experience were a single color, one could not know which color it was—or that it was a color at all. Our method of identification is structural and holistic. Let us now turn to the implications of this method for NCC research specifically.

4 From structuralism to structural constraints on NCCs

We began by pointing out that in psychophysics we capture perceptual experiences relationally in terms of our ability to match and differentiate them, even if the difference is only barely noticeable. We thereby construct phenomenal spaces that represent how similar or different any two basic experiences are, for instance the CIE L*a*b* color space (see section 4.3).

The fact that phenomenal experiences hang together in different degrees of similarity and difference is uncontroversial—it is common ground even with qualia

²³For further details on the empirical claims of this paragraph, see Hardin (2014).

enthusiasts who consider qualitative feels to be additional, intrinsic properties that are not individuated by the structure they exist within. The qualia enthusiast may even concede that location within a structure may uniquely identify an experience, but not that this makes an experience what it is (Levine, 1995). Instead, qualia are supposed to be something *additional* to this structure.

In scientific contexts, it suffices if we can uniquely identify basic experiences by their location in a structure. Beyond this, scientists may be agnostic as to whether there is anything that individuates experiences other than their location in a structure. Still, the type of structuralism employed in psychophysics has interesting methodological ramifications for how we investigate NCCs and which neural areas count as candidates for NCCs proper.

4.1 Structural constraints on experiments in NCC research

The search for NCCs is at the heart of contemporary consciousness studies. The NCC of an experience c is understood as the neural state s that is minimally sufficient for a c -experience to occur. But if each experience can be determined only by its relations to all other experiences, we cannot search for the NCC of a single experience *by itself*: An isolated neural state can never be sufficient for a specific experience. Without relational context, we cannot say what experience a neural state is a correlate of.

Therefore, we cannot search for the neural correlate of an individual color like Red_{42} . Instead we first have to search for the neural correlate of the color space as a whole, i.e. the neural system that can be in states where some are minimally sufficient for each of the perceivable colors. But we have to consider the whole range of states of this neural system, as these are necessary to make any individual state minimally sufficient for *this* experience. Only the neural correlate of the color space as a whole allows us to determine which neural state is the NCC for a specific experience like Red_{42} .²⁴

This leads to two methodological suggestions for experiments that deal with stimuli to elicit experiences:

- (i) **Intra-individual holistic mapping:** In order to specify which conscious state we are finding a correlate for when we record an individual's brain activity elicited by a specific stimulus (e.g., a photon of 550 nm), we must first psychophysically map out which *other* perceptual experiences are possible for that individual. Only this holistic mapping of experiences allows us to determine which experience is attached *to this stimulus*.

²⁴Here we call neural systems the NCCs of a group of phenomenal types. But they are NCCs only derivatively, due to the states they can be in, where some of their neural states correspond to types of experiences. Whenever such a state is instantiated, this is the NCC of a token of that phenomenal type.

(ii) **Inter-individual difference mapping:** We cannot rule out the possibility that other people may experience more or less than we do, e.g., that di- or tetrachromats may experience fewer or more shades of color. Because (as phenomenal holism states) each phenomenal experience can be individuated only by its relations to *all* other experiences, this means that such individuals must also have *different* experiences than trichromats do, thanks to the differences in the structure of their nervous system. Therefore, we have to capture inter-individual differences because only then can we determine which experience is attached to a specific stimulus *in this specific individual*.

Only if we map holistically can we pool across individuals and ensure that we find a more general neural correlate for a specific type of experience like Red₄₂.

4.2 Structural constraints on NCCs proper

We began this paper with the question of how to distinguish mere statistical correlates from the NCC proper. Here, neurophenomenal structuralism is helpful.

First, let us stress that the second assumption of neurophenomenal structuralism—that phenomenal structure is mirrored in certain neural structures—is an undercurrent in contemporary neuroscience and supported by some empirical evidence (see also section 4.3). For example, the geometry of the visual field maps onto the geometry of the visual cortex (Klein et al., 2004). Based on such findings, inter-individual differences in experiences can be predicted by inter-individual neural differences. For instance, differences in the size of V1 (Genç et al., 2015; Schwarzkopf et al., 2011) or the parietal cortex (Kanai et al., 2010) allow us to predict how long it will take for images to switch in a binocular rivalry paradigm for a single individual. This is likely not only to hold for the internal structure of individual experiential tokens (like visual impressions), but also to generalize to the structures governing phenomenal types (like colors). We argued in section 3 that due to the mere-sufficiency constraint of the NCC—according to Chalmers (2000), a neural state is seen as merely sufficient for a phenomenal experience, and not necessary—this relation will take the form of a surjective homomorphism of the neural onto the phenomenal.²⁵

If the relation between the neural and the phenomenal is generally systematic, we can expect that the neural correlate of a holistically structured phenomenal space should exhibit the structural features of this space. Hence, the NCCs proper of a phenomenal domain must somehow possess a structure corresponding to the relevant phenomenal space. The color-NCC proper must share the structure of the color space. If phenomenal state *a* is grouped more closely to *b* than to *c*,

²⁵Interestingly, Haynes (2009) has argued in a comparable fashion for a “homeomorphic (i.e. topology-preserving) mapping between a set of experiences and the states of their core NCC,” but focused on spatial patterns in and between objects, not basic experiences like colors.

then the neural states corresponding to *a*, *b*, and *c* must in some respect preserve this grouping.²⁶ We can expect to find structural similarities between phenomenal spaces and neural activation spaces: Neural data can be ordered in a space where the dimensions are the units of activation and a point in the spaces represents a particular pattern of activation code (Churchland & Sejnowski, 1992, p. 167 ff.). Within the neural activation space of an NCC, closeness and distance are indicators of degrees of similarity between the associated phenomenal experiences, even if the neural systems instantiating these states are not in anatomical proximity. One illustration of such a mapping can be found in (Churchland, 2005), where the structure of the Munsell spindle is mapped onto a subspace of the activation space of color-opponent processing cells in the Hurvich–James model. Proximity in this activation space is here reliant not on spatial proximity of the opponent processing cells, but on similarity of their respective activations.

Thus the neurophenomenal structuralist's approach suggests that structure is preserved if we cross over from phenomenal spaces to their neural substrates. Conversely, any neural system that lacks structural features of the relevant phenomenal space *fails* to be an NCC proper. It might statistically co-occur or contribute to proper functioning, but it cannot be the neural substrate of the phenomenal space. We call this the *structural similarity constraint* on NCCs (SSC):

(SSC) Neural substrates of phenomenal types must share the structure governing the phenomenal types they are associated with, i.e., the structure of a phenomenal space.

This, we believe, is an important constraint because it allows for a methodological refinement in empirical research: We can now distinguish mere statistical correlates from the NCC proper. Given that most structures patterning perception are highly complex and asymmetrical, we should not expect the same structure to be found in the state space of any arbitrary neural system. We can therefore use it to distinguish mere statistical correlates and artefacts from true neural substrates, since only the true neural substrates will capture all structural features patterning phenomenal experiences. Structuralism therefore gives us an exclusion criterion, differentiating merely statistical correlates from substrates in NCC research.

The distinction between substrates and statistical correlates might also apply to data collected by methods that are not temporally fine-grained. For example, even if a longer duration is captured in each “frame” of an fMRI recording, only part of the collected data-chunk will be maximally isomorphic to the structure of the phenomenal space. Neurophenomenal structuralism may therefore lead to an improvement in the methodology and tools used in NCC research. It may also lead to the rejection of several statistically correlating candidates as NCCs proper.

²⁶This will not always be as obvious as in the case of the visual field, where visual proximity maps largely onto proximity in V1. Some mappings might be much more counterintuitive.

These consequences are not mere speculation. Indeed, we see this type of reasoning as implicit in some existing empirical studies. In the following we present an illustrative example from contemporary neuroscience.

4.3 An example from neuroscience: Brouwer & Heeger (2009)

The idea that there are similarities between mental domains and their corresponding neural substrates is the second assumption of neurophenomenal structuralism (see section 3.3) and the core of the SSC. It is also a strong undercurrent in contemporary neuroscience. This idea is central to so-called representational geometry methods (Kriegeskorte & Kievit, 2013), where the assumption is that neural systems encode contents by integrating their (geometrical) structure into neural firing patterns. A research question guiding these approaches is: Do activation patterns in relevant neural areas stand in relation to each other in the same way that experiences do? For example, are neural activation patterns linked to perceiving red more similar to neural activations associated with orange than to patterns associated with green?²⁷ Cichy et al. (2019) applied representational similarity analysis to investigate the fit between behavioral similarity ratings of objects on different dimensions (color, shape, function, etc.) and fMRI as well as MEG activation space in visual cortex; they found neural encodings of psychophysical similarity spaces (for further neuroscientific perspectives see Haynes, 2009). The distance within such geometries is relative dissimilarity and their dimensions are defined by the specific aspects that are dissimilar. For our example here, representational geometry may measure perceived dissimilarities in the hues and relate it to neural dissimilarities such as how uncorrelated specific firing patterns are. In this way, neuroscientists (such as Cichy et al., 2019) analyze whether there is a structural similarity between perceptual and neural structure.

We argue that such a structural similarity constraint can be used to refine the concept of the NCC. The core idea is that if we knew the particular structure of experienced color dissimilarities, we would be able to use that structure to home in on the NCC proper of color experience because, according to the SSC, the NCC proper of any phenomenal domain must share the structure of the relevant phenomenal space. In the following, we focus on a study by Brouwer & Heeger (2009) which we see as illustrative of how the SSC may be used to distinguish mere statistical correlates of consciousness from NCCs proper.²⁸

{#CIE} Brouwer & Heeger (2009) focused on color experiences: They investigated the similarity relations between neural activation patterns linked to color stimuli in the visual cortex and compared them with a color space—specifically,

²⁷Note that activation pattern similarities might abstract from brain topology, since such patterns can be distributed over large areas.

²⁸We are not suggesting that the authors interpret their results in light of NCCs, but we think that their study could be used to do exactly that.

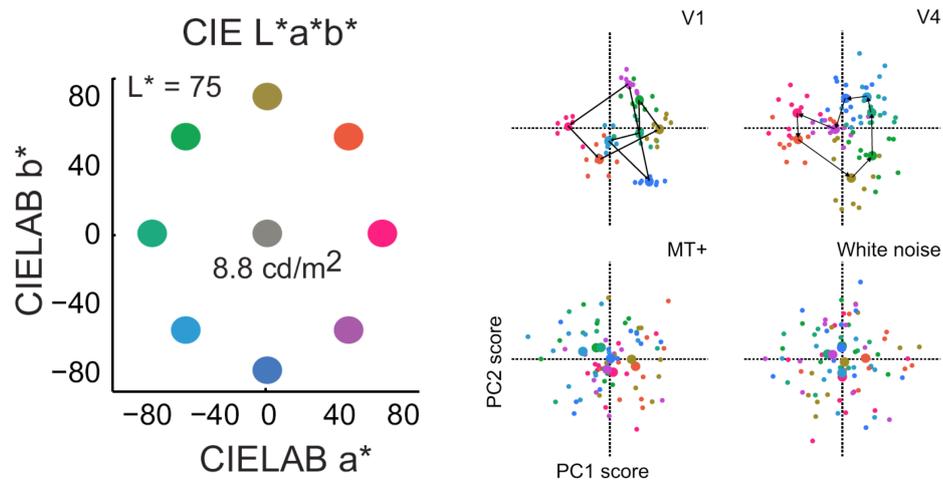


Figure 1: On the left side: The color stimuli used in Brouwer & Heeger (2009). On the right side: The result of the authors' principal component analysis clustering in V1, V4, and MT+ as well as white noise. The similarity in V4 manifests as a closed, non-overlapping loop, corresponding most significantly with the similarity structure of the section in the chosen color space. (Copyright 2009, Society for Neuroscience.)

a section from the CIE $L^*a^*b^*$ color space. This CIE color space is designed to be perceptually uniform, i.e., the distances in this space reflect perceived degrees of dissimilarity between colors (Commission Internationale de l'Éclairage, 1986). Brouwer and Heeger took a ring-shaped section from this space, consisting of eight individual color stimuli of equal lightness. These eight were chosen such that the similarity between each neighbor was equidistant in CIE $L^*a^*b^*$ color space. Additionally, each color was equidistant from neutral gray. Thus, the colors shown to the subjects possessed a circular similarity structure (see the left side of Fig. 1).

They then measured fMRI activation patterns for these eight colors. The fMRI activation patterns in the visual cortex that corresponded to the chosen colors were plotted into an underlying two-dimensional space.²⁹ Voxel activations related to the same colors were clustered. For V4, those clusters formed a closed shape. Thus, in V4, fMRI activation patterns associated with red were more similar to orange-related patterns than to green-related patterns, mirroring the structural organization of the associated experiential domain (see right side of Fig. 1).

Brouwer and Heeger's study offers an example of how researchers might use the structural similarity constraint (SSC) to distinguish statistical correlates from

²⁹This space was obtained by principal component analysis. Principal components are underlying dimensions that explain most of the variability within a dataset.

NCCs proper. According to the structural similarity constraint, some part of the NCC of color experience *must* encode the structure in the experiential domain as captured by psychophysics.

Brouwer and Heeger used principal component analysis to display neural patterns in a two dimensional neural space in order to extract similarity patterns from the recorded fMRI activation signals. According to this method, the activation of V4 displays a structure that resembles the specific hue circle which Brouwer and Heeger extracted from the CIE space. Other methods have already shown that V1 does process color but, in contrast to V4, V1 activation does not show the circular, roughly equidistant structure prominent in experiences elicited by the color stimuli. Region MT+ exhibited only randomly distributed patterns. So while V4 preserves the closed, non-overlapping loop-structure of the hue circle, V1 fails to show this structure.³⁰

From the perspective of our SSC criterion, we can interpret the above findings in the following way: Even though its activation statistically correlates with color experiences, V1 is likely not an NCC proper for color experience because it orders colors in a different way than we experience them. V1 does process color, but likely in a phase preceding color experience. MT+, on the other hand, does not process color in any way. Thus, within the visual cortex (and in the context of fMRI), the SSC favors V4 as a candidate for an NCC proper of color experience. Compared to other areas, it best encodes the structure of a specific phenomenal color space in terms of its circular form and clustering. Based on the SSC, we take the study by Brouwer & Heeger (2009) to indicate that V4 is the currently best candidate for an NCC proper of color experience, while other areas such as V1 are likely mere statistical correlates.

4.4 Limitations and applications

Can the SSC be applied outside of specific experiments and contribute to broad-scale theoretical discussions? To a degree.

The SSC may usefully contribute to the debate around frontal versus non-frontal theories of consciousness, where the question is whether NCCs extend to the non-sensory areas of the brain, especially the frontal cortex. Frontalists claim that they must—in part because the projection of information into the frontal areas may facilitate the formation of higher-order thoughts, which are supposed to make mental qualities conscious. Non-frontalists argue that this conflates phenomenal consciousness with access consciousness, which is all that is facilitated by the projection to the frontal areas.

Frontalists who accept the SSC must expect the prefrontal cortex, a region linked to higher cognitive functions, to *re-encode* the structural color information available in V4. This is certainly possible. If it does not, the prefrontal cortex will

³⁰Interestingly, V1 showed better decoding results than V4. This might indicate that mere information processing diverges from structural content.

not count as an NCC proper according to the SSC. Still, it could be that prefrontal cortex activation is needed for consciousness even though it might *not* encode that very structure (Haynes, 2009). For instance, it could function as an enabling condition. One interpretation available to frontalists in this scenario would be that the structure-preserving NCC proper of colors in the sensory areas is a kind of content condition, determining what specifically we experience. This would make it similar to the neural correlate of content (NCCc) which Marvan and Polák (2020) suggest needs to be added to a general neural correlate of consciousness (gNCC).

How decisive is the SSC? Again, the SSC is only a comparative constraint: V4 does not perfectly capture the structure of the color space used by Brouwer and Heeger. Still, compared to the other areas, it captures the similarity structure of the experience *best*. Therefore, V4 is the currently best candidate for an NCC proper.

This comparison by SSC may, however, reach a limit without yielding one and only one NCC proper. Neural areas other than V4 might fulfill the SSC for color equally well if compared against V4. We could then not use SSC to decide whether this hypothetical area or V4 is *the* NCC of color consciousness. But this is to be expected: The classical NCC definition explicitly allows that several neural areas can bring about the same experience. Still, according to the SSC, any area that counts as minimally sufficient would need to encode the corresponding phenomenal structure. The SSC is thus a feature that needs to be shared by all individual NCCs, each of which is only minimally sufficient. This aligns with the concept of the NCC 2.0 proposed by Fink (2016): Any token NCC must share features that make it an NCC of a specific type of experience. Manifesting structural similarities with the relevant phenomenal spaces might just be such a feature.

What are limitations of the SSC? Admittedly, there are some, and they are mainly imposed by the granularity and reliability of the methods we currently have available. First, there are limitations on the degree to which psychophysics captures the one true structure of phenomenal experiences. As a matter of fact, different psychophysical methods lead to phenomenal spaces with slightly different structures (see Kuehni, 2003 for a survey on color spaces).³¹ All of them capture some feature of our experience. Which is the right one? This will be hard to determine and more foundational theoretical work is needed. Maybe phenomenal structure is so simple that only those features that are invariant across methods truly grasp phenomenal structures, while others are artefacts. Maybe phenomenal structure is so complex that each psychophysical method captures only part of it. Maybe phenomenal structure is fluid and shifting such that there is no unique structure. Currently, it is too early to tell. At the moment, we can only say that insofar as psychophysics captures some structures of experience, these structures must be shared by the relevant NCC. Methodologically, the SSC thus makes NCCs

³¹Brouwer & Heeger (2013) were able to replicate their initial findings with another color space, the DKL space (Derrington et al., 1984). Again, V4 delivered the best match compared to other visual areas like V1, V2, and V3. This suggests that there is some transferability of these results to other color space models.

relative to a method for capturing phenomenal structures. Unless we have methods that allow us to distinguish psychophysical artefacts from those structures that are adequately captured, there is hardly a way to avoid this. But this is a general point that applies to all methods capturing first-person phenomena, not only to the SSC: All might produce artefacts.

A second limitation of the SSC concerns the methods for measuring neural activation. If these are not fine-grained enough, we may miss structural similarities, for instance on the level of individual neurons rather than columns. Or we might see similarities with phenomenal spaces in neural activation patterns that are only artefacts. Here, we can only hope for further progress in brain measurement techniques. Without it, the SSC may not prove decisive for a large swath of NCCs. Additionally, earlier exclusions based on the SSC may need to be revised in the light of new methodological approaches. But this is not uncommon in neuroscience.

Third, taking the two previous points together, if phenomenal structures are captured only in a very simplified fashion (for instance, if the mere cardinality of distinguishable experiences is captured, but not relative similarity or dissimilarity), we might find structural similarities in any area of the brain. The SSC would then never be decisive. This may be the case for some very simple sensory modalities, like our feeling for up versus down or hunger versus satiety. However, for most experiences, and especially colors, the relevant psychophysical spaces show a lot of complexity and asymmetry, which will hardly be preserved by all areas in the brain. Again, improvements in methods for capturing phenomenal structures as well as brain structures may make the SSC more and more decisive over time.

In summary, we claim simply that the SSC helps to limit NCC candidate areas comparatively. The SSC is not decisive in and of itself, but adds to other constraints that aid us in selecting the areas that serve as true neural correlates of a certain kind of *specific* experience. The SSC does not pick out a single NCC proper, nor does it pinpoint a general mechanism for all kinds of consciousness, but it does identify specific phenomenal domains with certain structures of sufficient complexity. And here, again, the SSC is constrained by the granularity of our methodology. Still, we hope to have illustrated via our reinterpretation of Brouwer & Heeger's (2009) findings that the SSC is valuable because it constrains what can count as an NCC proper.

5 Conclusion

Experiences stand in complex relations to each other, which are captured by psychophysics using a variety of methods. In contrast to qualia intrinsicism, phenomenal structuralism claims that it is this net of relations, the structure governing the types of experience we can have, that helps us to individuate any single experience. This entails a methodological holism, because we need to know the structure as a whole in order to determine the nature of the relata.

Neurophenomenal structuralism comes with the additional thesis that these phenomenal relations are mirrored in neural structures. Motivated by the empirically inspired mere-sufficiency constraint of Chalmers' NCC, we expect a surjective homomorphism of the neural onto the phenomenal. Structuralism is very much in tune with empirical methods like representational similarity analysis and the close historical lineage of NCC research to inner psychophysics, and is a good fit for many empirical findings.

Finally, neurophenomenal structuralism motivates a structural similarity constraint (SSC) that can contribute to NCC research. Since phenomenal structure must be mirrored in neural structure, we should exclude as NCCs proper all neural activation in areas that fails to capture the relevant phenomenal space structures. This structural similarity constraint on NCCs is, we believe, an important step towards reliably distinguishing neural substrates from statistical noise in neural data.

Acknowledgments

We are grateful for the comments of two anonymous reviewers. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 337619223 / RTG2386. We also want to thank Emily Troscianko for her editorial support.

References

- Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews*, 36(2), 737–746. <https://doi.org/10.1016/j.neubiorev.2011.12.003>
- Billock, V. A., Gleason, G. A., & Tsou, B. H. (2001). Perception of forbidden colors in retinally stabilized equiluminant images: An indication of softwired cortical color opponency? *Journal of the Optical Society of America*, 18(10), 2398–2403. <https://doi.org/10.1364/josaa.18.002398>
- Block, N. (1998). Semantics, conceptual role. In E. Craig (Ed.) *Routledge encyclopedia of philosophy*. Routledge. <https://doi.org/10.4324/9780415249126-W037-1>
- Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *The Journal of Neuroscience*, 29(44), 13992–14003. <https://doi.org/10.1523/JNEUROSCI.3577-09.2009>
- Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *The Journal of Neuroscience*, 33(39), 15454–15465. <https://doi.org/10.1523/JNEUROSCI.2472-13.2013>
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness* (pp. 17–39). The MIT Press.
- Chirimuuta, M. (2015). *Outside color: Perceptual science and the puzzle of color in philosophy*. The MIT Press.
- Churchland, P. (2005). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology*, 18(5), 527–560. <https://doi.org/10.1080/09515080500264115>
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. The MIT Press.
- 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. (2000). *A theory of sentience*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198238515.001.0001>
- Commission Internationale de l'Eclairage. (1986). *CIE 15.2, Colorimetry*. (2nd ed.) Commission Internationale de l'Eclairage.
- Crick, F. (1995). *The astonishing hypothesis: The scientific search for the soul*. Scribner.
- Decock, L. (2006). A physicalist reinterpretation of 'phenomenal' spaces. *Phenomenology and the Cognitive Sciences*, 5(2), 197–225. <https://doi.org/10.1007/s11097-005-9006-7>

Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on neural correlates of consciousness *Philosophy and the Mind Sciences*, 2, 7. <https://doi.org/10.33735/phimisci.2021.79>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369

- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, 357, 241–265. <https://doi.org/10.1113/jphysiol.1984.sp015499>
- Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel.
- Fink, S. B. (2016). A deeper look at the "neural correlate of consciousness". *Frontiers in Psychology*, 7, 1044. <https://doi.org/10.3389/fpsyg.2016.01044>
- Fink, S. B. (2018). Introspective disputes deflated: The case for phenomenal variation. *Philosophical Studies*, 175(12), 3165–3194. <https://doi.org/10.1007/s11098-017-1000-8>
- Fodor, J. A., & Lepore, E. (1992). *Holism: A shopper's guide*. Blackwell.
- Genç, E., Bergmann, J., Singer, W., & Kohler, A. (2015). Surface area of early visual cortex predicts individual speed of traveling waves during binocular rivalry. *Cerebral Cortex*, 25(6), 1499–1508. <https://doi.org/10.1093/cercor/bht342>
- Gert, J. (2017). Quality spaces: Mental and physical. *Philosophical Psychology*, 30(5), 525–544. <https://doi.org/10.1080/09515089.2017.1295303>
- Gescheider, G. A. (1997). *Psychophysics: The fundamentals*. (3rd ed.) Routledge.
- Goff, P. (2017). *Do electrons dream of electric sheep?* <https://blogs.lse.ac.uk/theforum/go%0B-do-electrons-dream/>
- Hardin, C. L. (2014). More color science for philosophers. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities* (pp. 379–390). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199832798.003.0016>
- Haynes, J.-D. (2009). Decoding visual consciousness from human brain signals. *Trends in Cognitive Sciences*, 13(5), 194–202. <https://doi.org/10.1016/j.tics.2009.02.004>
- Hofer, H., Singer, B., & Williams, D. R. (2005). Different sensations from cones with the same photopigment. *Journal of Vision*, 5(5), 444–454. <https://doi.org/10.1167/5.5.5>
- Hohwy, J. (2011). Phenomenal variability and introspective reliability. *Mind & Language*, 26(3), 261–286. <https://doi.org/10.1111/j.1468-0017.2011.01418.x>
- Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, 1(II). <https://doi.org/10.33735/phimisci.2020.II.64>
- Isaac, A. M. C. (2014). Structural realism for secondary qualities. *Erkenntnis*, 79(3), 481–510. <https://doi.org/10.1007/s10670-013-9519-3>
- Isaac, A. M. C. (2018). Prospects for timbre physicalism. *Philosophical Studies*, 175(2), 503–529. <https://doi.org/10.1007/s11098-017-0880-y>
- Jameson, K. A., Highnote, S. M., & Wasserman, L. M. (2001). Richer color experience in observers with multiple photopigment opsin genes. *Psychonomic Bulletin & Review*, 8(2), 244–261. <https://doi.org/10.3758/bf03196159>
- Kanai, R., Bahrami, B., & Rees, G. (2010). Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current Biology*, 20(18), 1626–1630. <https://doi.org/10.1016/j.cub.2010.07.027>
- Klein, I., Dubois, J., Mangin, J.-F., Kherif, F., Flandin, G., Poline, J.-B., Denis, M., Kosslyn, S. M., & Le Bihan, D. (2004). Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Brain Research. Cognitive Brain Research*, 22(1), 26–31. <https://doi.org/10.1016/j.cogbrainres.2004.07.006>
- 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>
- Kuehni, R. G. (2003). *Color space and its divisions: Color order from antiquity to the present*. John Wiley & Sons.
- Kuehni, R. G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research & Application*, 29(2), 158–162. <https://doi.org/10.1002/col.10237>
- Kuehni, R. G., & Schwarz, A. (2008). *Color ordered: A survey of color order systems from antiquity to the present*. Oxford University Press.
- Lawless, H. T. (2013). *Quantitative sensory analysis: Psychophysics, models and intelligent design*. Wiley Blackwell.
- Levine, J. (1995). Qualia: Intrinsic, relational, or what? In T. Metzinger (Ed.), *Conscious experience* (pp. 277–292). Ferdinand Schöningh.
- Lewis, C. I. (1990). *Mind and the world-order: Outline of a theory of knowledge* (Unabridged republication of the 1929 edition). Dover Publications.
- Marvan, T., & Polák, M. (2020). Generality and content-specificity in the study of the neural correlates of perceptual consciousness. *Philosophy and the Mind Sciences*, 1(II). <https://doi.org/10.33735/phimisci.2020.II.61>
- Niederée, R. (2010). More than three dimensions: What continuity considerations can tell us about perceived color. In J. Cohen & M. Matthen (Eds.), *Color ontology and color science*. The MIT Press. <https://doi.org/10.7551/mitpress/7541.003.0006>
- O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines & P. Slezak (Eds.), *Representation in mind* (pp. 1–20). Elsevier. <https://doi.org/10.1016/B978-008044394-2/50004-X>

Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on neural correlates of consciousness *Philosophy and the Mind Sciences*, 2, 7. <https://doi.org/10.33735/phimisci.2021.79>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369

- Papineau, D. (2015). Can we really see a million colours? In P. Coates & S. Coleman (Eds.), *Phenomenal qualities* (pp. 274–297). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198712718.003.0010>
- Pearson, K. (1897). Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London*, 60, 489–498. <https://doi.org/10.1098/rspl.1896.0076>
- Rosenthal, D. (2015). Quality spaces and sensory modalities. *Phenomenal Qualities: Sense, Perception, and Consciousness*, 33–65.
- Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, 14(1), 28–30. <https://doi.org/10.1038/nn.2706>
- Seth, A. (2009). Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*, 1(1), 50–63. <https://doi.org/10.1007/s12559-009-9007-x>
- Seth, A. K., & Edelman, G. M. (2009). Consciousness and complexity. In R. A. Meyers (Ed.), *Encyclopedia of complexity and systems science* (pp. 1424–1443). Springer New York.
- Shepard, R. N., & Chipman, S. (1970). Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1(1), 1–17. [https://doi.org/10.1016/0010-0285\(70\)90002-2](https://doi.org/10.1016/0010-0285(70)90002-2)
- Tsuchiya, N., Phillips, S., & Saigo, H. (2021). *Enriched category as a model of qualia structure based on similarity judgements*. OSF Preprints. <https://doi.org/10.31219/osf.io/5peu7>
- Wiese, W. (2018). Toward a mature science of consciousness. *Frontiers in Psychology*, 9, 693. <https://doi.org/10.3389/fpsyg.2018.00693>
- Young, B. D., Keller, A., & Rosenthal, D. (2014). Quality-space theory in olfaction. *Frontiers in Psychology*, 5, 1. <https://doi.org/10.3389/fpsyg.2014.00001>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

