Psychoneural Isomorphism

From Metaphysics to Robustness

Alfredo Vernazzani

Institut für Philosophie II, Ruhr-Universität Bochum

Abstract At the beginning of the 20th century, Gestalt psychologists put forward the concept of psychoneural isomorphism, which was meant to replace Fechner's obscure notion of psychophysical parallelism and provide a heuristics that may facilitate the search for the neural correlates of the mind. However, the concept has generated much confusion in the debate, and today its role is still unclear. In this contribution, I will attempt a little conceptual spadework in clarifying the concept of psychoneural isomorphism, focusing exclusively on conscious visual perceptual experience and its neural correlates. Firstly, I will outline the history of our concept, and its alleged metaphysical and epistemic roles. Then, I will clarify the nature of isomorphism and rule out its metaphysical role. Finally, I will review some epistemic roles of our concept, zooming in on the work of Jean Petitot, and suggest that it does not play a relevant heuristic role. I conclude suggesting that psychoneural isomorphism might be an indicator of robustness for certain mathematical descriptions of perceptual content.

Acknowledgments For financial support, I would like to thank the Barbara-Wengeler-Stiftung and the Volkswagenstiftung's project "Situated Cognition. Perceiving the World and Understanding other Minds" led by Tobias Schlicht. For comments on earlier versions, many thanks also to Marcin Miłkowski, Albert Newen, Marco Viola, Fabrizio Calzavarini, Krys Dołega, Judith Martens, Elmarie Venter, Luke Roelofs, Antonio Piccolomini d'Aragona, Francesco Marchi, Brendan Ritchie, and Franceso Altiero.

1. Introduction

At the beginning of the 20th century, the Gestalt psychologists put forward the concept of *psychoneural isomorphism* (Köhler 1929), the claim that the "mind" and the "neural" are isomorphic. The Gestaltists' aim, as we will see, was that of replacing the vague concept of psychophysical parallelism that constituted the philosophical foundation of much of early 19th century psychophysics. Yet, the concept has never been fully clarified and in contemporary contributions it still represents a source of puzzlement[1]. It is far from clear in what sense the mental and the neural would be isomorphic, or what theoretical purpose the concept is supposed to play.

My purpose is to provide a conceptual roadmap of psychoneural isomorphism, one that can be used to dispel some misunderstandings and help the reader to frame the concept in the correct way, identifying the alleged roles of isomorphism with particular reference to contemporary debates.

In §2, I briefly reconstruct the history of our concept, locate it in the contemporary debate, and highlight its alleged roles. In §3, I make some conceptual clarifications, specifying what an isomorphism is and under what conditions we can properly speak of isomorphism. After distinguishing between an ontic and an epistemic reading, I turn to the ontic reading in §4, and on the epistemic reading in §5, taking as case-study Petitot's morphodynamical models. My contention is that while isomorphism arguably does not play the roles it has been traditionally associated with, there is an overlooked possibility that isomorphism might play a role in robustness analysis.

2. Why Psychoneural Isomorphism?

2.1. Historical Overview

The concept of psychoneural isomorphism was introduced in response to the 19th century debate about the philosophical foundations of psychology and psychophysics. In order to clarify our concept, it will be helpful to briefly sketch out the ideas of some key figures (§2.1.1) in response to which Köhler (§2.1.2) introduced the concept of isomorphism.

2.1.1 *From Fechner to Müller*. Over the course of his career in research, Gustav Fechner, the father of modern psychophysics, held different views. Early, in his dissertation *Praemissae ad theoriam organismi generalem*, he stated that «parallelismus strictus existit inter animam et corpus, ita ut ex uno, rite cognito, alterum construi possit» (quoted from Heidelberger 2000, p. 53) [A tight parallelism holds between soul and body, such that from one, properly understood, the other one may be constructed][2]. This proposition anticipates his core metaphysical view, the "identity perspective" (*Identitätsansicht*) according to which the soul and the body are but aspects of the same substance[3]. This view was further articulated

---

[1] Some contemporary contributions include: Bridgeman (1983), Lehar (1999, 2003), Noë & Thompson (2004), O'Regan (1992), Palmer (1999).

[2] All translations in this chapter are mine.

[3] Fechner's identity view owns a debt to Schelling, whose work exerted an influence on Fechner via Lorenz Oken's lectures, a *Naturphilosoph* whose ideas had been significantly inspired by the philosopher from Leonberg.

over the years. If in 1851 Fechner could still say that the soul's and the body's processes are «im Grunde nur dieselben Processe» [basically the same processes], later Fechner drew closer to an objective idealism (*objektiver Idealismus*) according to which the fundamental layer of reality is spiritual. It was, however, ultimately his philosophical commitment to the deep unity of soul (or mind) and body that led him to articulate a mathematical approach to psychophysics.

Fechner had a considerable influence over his contemporaries and the younger generation of psychologists in Germany. With the exception of Helmholtz and his followers, who espoused a form of dualism, most psychologists adopted the *Identitätsansicht* as a heuristic method, i.e. as a conceptual bridge that might help investigate the brain, or sought to replace it with better conceptualizations. Mach (1865) belonged to this second group of researchers. Inspired by Fechner, he formulated a "principle of equivalence" [*Princip der Entsprechung*], according to which for every psychological event there must be a corresponding physical event and that identical psychological events must correspond to identical physical events. In the revised 1900 edition of his *Analyse der Empfindungen* (1886) he argued that the «guiding principle for the study of sensations» [*leitender Grundsatz für die Untersuchung der Empfindungen*] would be the «principle of complete parallelism of the psychical and the physical» [*Princip des vollständigen Parallelismus des Psychischen und Physischen*]. Mach meant this to be a «heuristic principle» [*heuristisches Princip*], which constitutes the «necessary presupposition of exact science» [*notwendige Voraussetzung der exakten Forschung*; 1886/1922, p. 50; in Scheerer 1994, p. 320]. Later, in the 1906 edition of his magnum opus, Mach stated to be looking for «similarity of form» [*Formähnlichkeit*] between the physical and the psychical and vice-versa (Heidelberger 2000).

In the pages of his work dedicated to the forerunners of psychoneural isomorphism, Köhler (1929) did not discuss Mach's ideas—this is oddly enough, since the development of Gestalt psychology was heavily influenced by Mach and Ehrenfels (Greenwood 2015, pp. 326-327). Köhler, however, discussed Hering's principle of parallelism, and most importantly, George Müller's psychophysical axioms (1896). Müller was Friedrich Schumann's teacher and mentor, who later became Carl Stumpf's collaborator and one of Wertheimer's teachers in Berlin. Schumann later moved to Frankfurt, where he hired as assistants both Köhler and Koffka. A fervent admirer of Fechner, Müller put forward five axioms with the explicit purpose of replacing the notion of psychophysical parallelism and provide a better heuristic principle (1896, p. 4). It is not possible to discuss the axioms in detail here, it suffices to foreground that the first three axioms established a correspondence between mental, conscious states and their variations with underlying psychophysical processes.

2.1.2 Köhler. In his 1929 book, Köhler launched an attack against Behaviorism and pointed up the importance of first-person approaches to the study of the mind. According to Köhler, psychologists should investigate the «terra incognita» that lies between sensory stimulation and overt behavior:

> To the degree to which the interior of the living system is not yet accessible to observation, it will be our task to invent hypotheses about the events which here take place. For much is bound to happen between stimulation and response. (Köhler 1929, p. 51).

Köhler was aware of the limitations of early 20th century neuroscience, and devised a principle that, exploiting the dependence relation of the mind upon brain processes (ivi, p. 57), could be used to infer something about the latter given that the «[e]xperienced order in space is always structurally identical with a functional order in the distribution of the underlying brain processes», he called this principle «*psychophysical isomorphism*» (ivi, pp. 61-62). Although Köhler did not clearly define psychophysical (or "psychoneural", as it came to be called) isomorphism, he clearly understood this as a contribution to the debate sparkled by Fechner's *Identitätsansicht*.

Indeed, Köhler's terminology is often unclear. For example, in his 1938 book, he defined psychoneural isomorphism a helpful "postulate" in the formulation of empirical hypotheses (Luccio 2010, p. 228). The terminology is unfortunate, for a postulate is a proposition that is simply assumed as true, but in several passages, Köhler seemed less committed towards isomorphism. In a late work, Köhler described the principle not as an a priori postulate, but as «an hypothesis which has to undergo one empirical test after the other» (quoted in Scheerer 1994, p. 188). But Köhler also persistently confused metaphysics with heuristic assumptions, and in later works, he seemed to embrace psychoneural isomorphism for the sake of a monistic metaphysics:

> For instance, if the comparison were to show that, say, in perception, brain processes with a certain functional structure give rise to psychological facts with a different structure, such a discrepancy would prove that the mental world reacts to those brain processes as a realm with properties of its own—and this would mean dualism (Köhler 1960, quoted in Luccio 2010, p. 241; my emphasis).
>
> …[monism] would become sensible precisely to the extent that isomorphism can be shown to constitute scientific truth. (Köhler 1960, quoted in Scheerer 1994, p. 189).

We will later (§3) examine whether psychoneural isomorphism supports a monistic account of the mind-brain relation.

2.2. *Psychoneural Isomorphism Today*

Glossing over some further developments both within and beyond Gestalt psychology[4], I want to draw the reader's attention to two recent debates in which our

---

[4] Noteworthy is the development of second-order isomorphism that would hold between «(a) the relation among alternative external objects, and (b) the relations among their corresponding internal representations» (Shepard & Chipman 1970, p. 2). More recently, second-order isomorphisms have been exploited in research on dissimilarity matrices among internal representations (e.g. Kriegeskorte et al. 2008; Kriegeskorte & Kievit 2013). As Kriegeskorte et al. (2008, p. 4) comment, this approach is «complementary» to that of «first-order isomorphism», i.e. psychoneural isomorphism. In this contribution I will exclusively focus on psychoneural isomorphism and leave an exploration of the relations between first- and second order isomorphisms as an open avenue for future research.

concept seems to be playing an important role: the problem of Naturalized Phenomenology (§2.2.1), and the search for neural correlates of consciousness (§2.2.2.).

2.2.1 Naturalized Phenomenology. Proponents of naturalized Phenomenology or Neurophenomenology argue that rigorous descriptions of our lived experience should complement the third-personal standpoints of the sciences of the mind since, it is argued, purely third-personal approaches cannot capture the nature of consciousness or conscious content. These rigorous descriptions should be produced by subjects trained in Husserlian Phenomenology—with capitalized "P" to distinguish it from our conscious experience, or "phenomenology"— (Roy et al. 1999; Varela 1997).

The need for conceptually rigorous descriptions is motivated by the unreliability of naïve introspective reports (Schwitzgebel 2011; Vernazzani 2016). Rigorous phenomenological descriptions, so understood, would offer a precious help in theory construction as well as theory confirmation (Roy et al., p. 12), identifying the neural structures that would explain our conscious experience. In the current debate, these structures are known as the neural correlates of consciousness or NCCs (Chalmers 2000; Wu 2018). In order to shed light on the nature of the biological correlates of conscious experience, however, phenomenological descriptions must be naturalized. By "naturalized" they mean «integrated into an explanatory framework where every acceptable property is made continuous with the properties admitted by the natural sciences» (Roy et al. 1999, pp. 1-2). According to Roy et al., the naturalization can be achieved by means of mutual constraints in theory construction (cf. also Flanagan 1992, p. 11; Varela 1997). In turn, mutual constraints are made possible either by means of psychoneural isomorphism, or "generative passages" (Roy et al. 1999, p. 66-68). The former concept is quickly and obscurely dismissed, as Roy et al. suggest that «this isomorphic option makes the implicit assumption of keeping disciplinary boundaries» and urge that psychoneural isomorphism would entail some form of «psycho-neural identity» (ivi, p. 68). The latter concept makes reference to identification of «passages» that one could in principle mathematically specify to allow a transition from the "phenomenological" to the "neural." As Bayne (2004) rightly points out, however, it is far from clear whether such generative passages are substantially different, and in what sense, from a psychoneural isomorphism.

2.2.2. The Search for Content-NCCs. The identification of neural structures or correlates related to conscious content (content-NCCs) represents a crucial goal in explaining our conscious perception. Focusing on conscious visual perception, an implicit assumption in much research is that there should be some sort of "matching" between experienced content and underlying neural representations in the content-NCCs (Crick & Koch 1995; 1998). Some researchers contend that the «proper form» (Pessoa et al. 1998, p. 726) of explanations in vision science necessarily involves talk about isomorphism between perceptual content and underlying neural processes (e.g. Fry 1948). Todorović argued that there must be an «identity of shapes of spatial distributions of percepts and the underlying neural activities» (1987, p. 548; cf. also Teller 1984).

Consider the case of the neon-color-spreading illusion, where a bright color seems to spread over a white background (cf. Bressan et al. 1997; Pessoa et al. 1998, pp. 730-731). An interesting question is whether it is a real phenomenon, or a conceptual artifact, on the basis of unreliable introspective reports (Dennett 1991; cf. §5.3). Assuming that it is a real phenomenon, the next question is whether a proper explanation of the neon-color spreading illusion (as well as other kinds of perceptual completion or filling-in; cf. Komatsu 2006; Pessoa & De Weerd 2003; Weil & Rees 2011) requires psychoneural isomorphism, i.e. a structural correspondence between percept and underlying neural activity. While some researchers opt for an isomorphic explanation (e.g. Von der Heydt et al. 2003, p. 107; Weil & Rees 2011, p. 41), not everyone agrees with this approach (e.g. Ratliff & Sirovich 1978), and Von der Heydt et al. contemplate an alternative explanation of filling-in of color by means of a tentative association of features by low-level mechanisms (symbolic filling-in theory).

The detailed explanation of filling-in phenomena is matter of empirical investigation, our concern here is whether, and in what sense, isomorphism plays an explanatory role, and whether we are justified in talking about an isomorphism in the first place. Noë & Thompson (2004) and Thompson (2007) have argued that there is no isomorphism between perceptual content and content-NCCs, the latter being understood as receptive fields of single neurons, since perceptual content exhibits properties—such as being intentional or "world-presenting," and holistic—that the latter lack. Petitot (2008, p. 396) rejects this conclusion on the ground that Noë and Thompson have wrongly identified single cells' receptive fields as the neural correlates of perceptual content, and argues that an emergent morphology of larger populations of neurons exhibits an isomorphism with perceptual content (§5.1).

2.3. The Roles of Psychoneural Isomorphism

While the debate on issues such as naturalized Phenomenology and the nature and search for content-NCCs is fraught with a great deal of thorny conceptual issues, we should keep the appropriate level of abstraction in order not to loose sight of our task, i.e. investigating psychoneural isomorphism. From the foregoing cursory overview, we can identify several roles that our concept is supposed to play:

- *Metaphysical role*. As we have seen, several researchers think that there must be a close connection between the metaphysics of the mind-brain, and psychoneural isomorphism. Köhler and Petitot for example think that isomorphism supports a monistic mind-brain metaphysics, i.e. a version of the identity theory (§2.1.2)[5]. I call the inference that isomorphism supports monism "From Iso-

---

[5] There are of course further complications. One complication is represented by the externalist challenge, i.e. whether the brain or the "neural" is the sole substrate of the mind. Another com-

morphism to Monism" (I-M) (§4.1). Others, like Revonsuo (2000), claim that if the mental and the neural are identical, then they must be isomorphic. I call this inference "From Monism to Isomorphism" (§4.2).

- *Heuristic Role*. Another important role ascribed to isomorphism is as heuristic principle. Much of the debate stirred by Fechner hinges on the search for better ways to articulate a heuristic principle that could help bridging the mind-brain gap at a time when neuroscience was still in its infancy (§2.1.1). The guiding idea seems to be the following. If we do not have any kind of access to *b*, but we have access to *a*, and we know that, under some level of description, *b* is isomorphic to *a*, then studying the structure of *a* will enable us to know the structure of *b*. It remains to be understood, however, whether, and to what extent, such a role might hold in the case of the mind-brain.
- *Explanatory Role*. One of neuroscientists' goals in searching for content-NCCs is explanatory (§2.2.2). As we have seen, it has been claimed that a content-NCC's neural representation must be isomorphic with the corresponding perceptual content. It is in virtue of this isomorphism that neurobiological models can be said to explain perceptual content (e.g. Noë & Thompson 2004; Pessoa et al. 1998; Roy et al 1999). Here, the term "model" is ambiguous. On an epistemic account of scientific explanations, models are epistemic representations that might be used to explain a given phenomenon (Wright 2012; §3.3). On the ontic account, it is the very thing itself, e.g. a worldly mechanism, which is said to explain the phenomenon (Craver 2014).
- *Intertheoretic Role*. Rigorous phenomenological descriptions of perceptual content have been said to be isomorphic with models or descriptions of the underlying neurobiological activity. This is at least the core idea behind proponents of naturalized phenomenology's mutual constraints, which is supposed to serve in «theory confirmation» and «theory construction» (Roy et al. 1999, p. 12). Framed in these terms, psychoneural isomorphism serves an important role in intertheoretic integration, i.e. the problem of integrating different fields or theories (e.g. Darden & Maull 1977).

These different roles should be sharply distinguished. It is noteworthy that the isomorphic relata in these roles are different kinds of entities that require substantive additional qualifications. On the metaphysical role we are talking about the mind itself, whereas in its intertheoretic role we are arguably talking about models of the mind. Accordingly, we can suggest the following partition between *ontic* and *epistemic* interpretations of psychoneural isomorphism (Tab. 1).

---

plication is the kind of identity theory assumed, tokens or types. As we shall see (§3.3), we can put these complications aside.

Ontic Isomorphism	Epistemic Isomorphism
Metaphysical Role Explanatory (ontic) role Heuristic role	Explanatory (epistemic) role Intertheoretic role Heuristic role

Table 1: The different roles of psychoneural isomorphism assume ontologically distinct relata. The ontic reading assumes that the isomorphic relata are worldly things, whereas the epistemic reading assumes the relata to be epistemic representations. The heuristic role can be placed on both sides.

In order to examine whether psychoneural isomorphism actually fulfills (some of) these roles, we must provide a robust definition of what an isomorphism is (§3.1), and then specify the nature of the relevant entities.

3. The Nature of Psychoneural Isomorphism

3.1. What Is An Isomorphism?

Unless the concept is used in a merely figurative way, the term "isomorphism" comes from mathematics[6]. More precisely, an isomorphism is a bijective homomorphism. A homomorphism is a function or map between two objects or domains that partially preserve their structures. Let us take two arbitrary domains *A* and *B* that are relational structures. A relational structure is a set *A* together with a family «R*i*» of relations on *A*. Two relational structures **A** and **B** are said to be similar if they have the same type. (I follow the convention of using a bold face **A** to refer to the relational structure, and the italics *A* to refer to the carrier set or domain). A homomorphism can be defined as follows:

> Let **A** and **B** be similar relational structures, with relations «R*i*» and «S*i*» respectively. A homomorphism from **A** to **B** is any function *m* from **A** into **B** satisfying the following condition, for each *i*: If $<a_1, a_2,\ldots a_n> \in Ri$, then $<m(a_1), m(a_2), \ldots m(a_n)> \in Si$. (Dunn & Hardegree 2001, p. 15)

B is the homomorphic image of **A** if there exists a homomorphism from **A** to **B**. Structural similarity admits different degrees. More informally, we can say that a homomorphic image **B** can be more or less structurally similar to **A**.

An isomorphism is a special kind of homomorphism. Thus, every isomorphism is also a homomorphism. For a homomorphism to be an isomorphism what is required is that the function *m* from **A** onto **B** must *completely* map the relational

---

[6] Brendan Ritchie has rightly pointed out to me that it is not clear whether Köhler understood isomorphism in the *mathematical* sense or rather in a more figurative and metaphorical sense. I have two replies to this. Firstly, although, as we have seen, Köhler did not give any rigorous definition, he meant this concept to provide a more rigorous and precise foundation than Fechner's notion of parallelism. This indicates that, arguably, he did not understood the concept as figurative or metaphorical. It should also be stressed that Köhler himself was certainly aware of the mathematical meaning of our concept, as he was trained in physics and mathematics under Max Planck. Secondly, and more importantly, the historical reception of our concept has clearly interpreted it in the mathematical sense (e.g. Madden 1957; Lehar 2003).

structure. In this sense, an isomorphism is a *bijective homomorphism*. A formal definition of isomorphism can now be given:

> A homomorphism *h* from **A** to **B** is said to be an isomorphism from **A** to **B** (between **A** and **B**) iff it satisfies the following conditions: (1) *h* is one-one; (2) *h* is onto. (Dunn & Hardegree 2001, p. 17)[7]

When a *h* satisfies these conditions, we may say for conciseness that **A** and **B** are *isomorphic*, or $\mathbf{A} \cong \mathbf{B}$. We can clarify our concept with the aid of an example. Consider the sequence of natural numbers $\mathbb{N}_0 = \{0, 1, 2, 3\ldots+\infty\}$. This sequence is isomorphic to the sequence of annual time segments from 0 to positive infinity, i.e. we can specify a function from the set of annual time segments to $\mathbb{N}_0$ that is homomorphic, one-one, and onto. Formally, as we have seen, every isomorphism is a special case of homomorphism, but for clarity's sake, I will use the term "homomorphism" for functions that by definition are less then isomorphic.

What are *A* and *B*? So far, I have construed the carrier sets as domains and isomorphism as a function between domains. But an isomorphism might hold also between topological spaces (a "homeomorphism"), rings, vector spaces, categories, etc. Furthermore, notice that one can also have a homomorphic function from **A** to **A**, i.e. when the domain and its image are identical. This is an interesting case. A *h* from **A** to **A** is called an *endomorphism*, i.e. a homomorphism from **A** to **A**. If *h* is one-one and onto we get an *automorphism*, i.e. an isomorphism from **A** onto **A** (Cohn 1981, p. 49). This is important because it shows that even if domain and image are identical, the function does not *have* to be an automorphism[8]. We can illustrate this by means of an example. Consider a vector space *V*, an endomorphism from **V** to **V** is a linear map:

$$L\colon V \to V$$

An automorphism is an invertible endomorphism (an invertible homomorphism). However, if we assume a vector dimension dim V > 0, the endomorphism $L\colon V \to V$ with $v \mapsto 0$ will *not* be invertible, hence, it is not an automorphism, although domain and image are identical.

To sum up, these are the jointly necessary and sufficient requirements for an isomorphism:

(1) We must identify a domain *A* and its image *B* (the carrier sets).
(2) We must show that *A* and *B* contain elements which stand in some relation with each other, i.e. that **A** and **B** are relational structures, and what kind of structures they are.

---

[7] Dunn and Hardegree define isomorphism by means of material implication. But usually, the concept is defined with a biconditional. I have rectified the quotation accordingly. Thanks to Christian Strasser for pointing this out to me.

[8] A limit case is the identity function, i.e. it is always possible to construct a function which returns the value used in the argument. Otherwise, however, the function must be specified (§4.2).

(3) We must identify a homomorphism h from $\mathbf{A}$ to $\mathbf{B}$ which is one-one and onto.

Talk of isomorphism that fails to meet these requirements can only be understood metaphorically and will not be discussed here.

3.2. What Does Psychoneural *Mean?*

Having clarified what an isomorphism is, we must now turn to its qualification as "psychoneural." The choice of domains of isomorphism is somewhat arbitrary, as it largely depends on our epistemic goals. The adjective "psychoneural" clearly suggests that our domain is the "psyche" or "mind" and the "neural" its image. Still, this leaves a great deal worth questioning.

Pribram (1983) stated that an isomorphism might hold between the brain and experience; or between the brain and the environment; or between all three. Arnheim argued that psychoneural isomorphism plays a fundamental role in conceptualizing the way we grasp other people's expressions (1949, pp. 58ff), with multiple domains being isomorphic. Madden (1957) distinguished between an isomorphism between stimuli and sensory responses; between receptor events and afferent neural processes; and between neural events and phenomenal (conscious) events. The latter pertains to what Fechner (1860) had called *internal psychophysics* (*innere Psychophysik*), i.e. the study of the relation between the brain and experience (*Erleben*). This is the isomorphism I will focus on, as it best captures Köhler's ideas, as well as the concept discussed in the contemporary debates, i.e. the neural correlates of consciousness and naturalized Phenomenology. Accordingly, I will have nothing to say about other putative isomorphisms, for example holding between retinal projection and the primary visual cortex V1[9].

There are some further clarifications in order. With reference to the discussions in §2.2.2 about naturalized Phenomenology as well as content-NCCs, it is clear that our concept is mostly discussed in relation to *conscious* visual *perceptual content*. With "consciousness" we shall understand the intrinsic or felt character of our mental lives, often characterized in terms of Nagel's construct "what-is-it-like-to-be" (1974). I shall occasionally refer to this aspect of consciousness as "phenomenology." This minimal characterization is neutral about whether there is an unbridgeable explanatory gap (Levine 1983) or whether consciousness might be reductively explained or ontologically reduced (e.g. Chalmers 1996). With "perceptual content" we shall understand, following the mainstream account of perceptual experience (e.g. Byrne 2001; Siegel 2010), the percipient's representational content at a given time[10]. I shall focus on "visual" perceptual contents, as the

---

[9] Similarly, I do not further discuss what we might call *implementational isomorphism*, i.e. the issue as of whether computational states are isomorphic to the underlying physical states or changes, for instance in the biological substrate (cf. Chalmers 2012; Miłkowski 2013; Piccinini 2015; Scheutz 2001).

[10] Different accounts of the nature of perception may impose different constraints on the domains. Within a naïve realist account (e.g. Brewer 2011), for instance, the locution "perceptual content" does not refer to representational contents, but to the very things we are directly percep-

studies I refer to (§2 and §4) zoom in on this particular modality, but my considerations might be easily extended to all other perceptual modalities as well. Not all perceptual contents are conscious, so our first domain is the domain of a subject S's conscious perceptual content at a given time *t*. I call this the *phenomenological domain*, Ψ[11]. It is widely assumed that our conscious mental lives depend (at least partly) on some subset of neural activity. I call this subset of neural activity from which the phenomenological domain depends, the *neural image* or *domain*, ϕ. I gloss over the nature of the neural domain, as this problem will be dealt with later (§4). I assume that the domains capture types, rather than token contents or neural structures.

Concerning the second requirement, it is assumed that the domains contain elements, and that such elements must stand in some relations with each others, i.e. these are not mere sets, but n-tuples of ordered elements. We can thus say that our domains carry, respectively, a *phenomenological relational structure* $\boldsymbol{\Psi}$ and a *neural relational structure* $\boldsymbol{\phi}$. Of course, one would have to specify such structures, but I will sidestep this issue for now. In general, determining, for instance, the nature of the phenomenological structure under examination will depend also by the specific nature of the domain and what sort of relations the elements in that domain might stand in. In order to satisfy the third requirement of isomorphism, we must specify a function *h* which is one-one and onto.

4. Psychoneural Isomorphism and the Metaphysics of the Mind

The foregoing considerations still leave open the question of the epistemic or ontic interpretation (§2.3). A moment reflection suggests that this distinction is not completely straightforward and requires further clarifications. After analyzing the alleged metaphysical roles of isomorphism (§§4.1-2), I will argue that the ontic reading is untenable (§4.3).

4.1. From Isomorphism to Monism

Earlier (§2.3), I identified two distinct metaphysical roles, the first was an inference "From Isomorphism to Monism" (I-M). The idea, roughly, is that if we can specify a psychoneural isomorphism, we thereby have some evidence for the identity of these domains. An instance of this strategy can be traced back to Köhler:

> …[monism] would become sensible precisely to the extent that isomorphism can be show to constitute scientific truth. (Köhler 1960, quoted in Scheerer 1994, p. 189; §2.1.2).

Another instance of this inference can be found in Petitot (2008) who, after showing that there is an isomorphism between morphological models *M* of the neurophysiology of the relevant functional architectures and morphological mod-

---

tually acquainted with. Accordingly, an isomorphism might hold between the observable aspects of things from a given standpoint, and the percipient's underlying neural activity.

[11] I assume a synchronic perspective, i.e. that of a subject ideally frozen at a time *t*. Alternatively, one could examine psychoneural isomorphism from a diachronic perspective, i.e. considering S's mental and neural states from t_1 to t_2.

els *E* of Husserlian descriptions of the phenomenal relation between experienced space and quality (§5.1), comments that this would warrant a double-aspect theory. A double-aspect theory, in the words of Metzinger, amounts to the following claim: «[s]cientifically describing and phenomenally experiencing are just two different ways of accessing one and the same underlying reality» (2000, p. 4). In other words, Petitot thinks that if there is an isomorphism between *M* and *E*, then brain activity and the phenomenal experience must be identical (monism).

Let us first make a preliminary clarification about the nature of identity. A distinction can be drawn between two kinds of identity (Noonan & Curtis 2014): qualitative and numerical. For two things to be qualitatively identical under some respect is for them to possess the same property. Max Ernst's *L'Ange du Foyer* and Paul Nash's *Totes Meer* both share the properties of "being a painting," "being surrealist artworks", etc. Qualitative identity may be spelled out in different ways, depending on our assumptions about the metaphysics of properties (cf. Allen 2016). It is clear that two entities may be qualitatively identical with respect to some, or most, properties, without they being one and the same thing. Numerical identity is much stronger. If *a* and *b* are numerically identical it means that *a* just is *b*. Numerical identity implies total qualitative identity. The mind-brain monism presently discussed is a debate about numerical identity, whether, ultimately, the mind just *is* the brain (or, better, some subset of its neural activity). With these clarifications, we can now throw light on the inference I-M. At first, one might think that we are dealing with something like this:

> (I-M)-1: If $\boldsymbol{\Psi} \cong \boldsymbol{\phi}$, i.e. there is an h, such that h is one-one and onto between the given relational structures, then $\boldsymbol{\Psi}$ is qualitatively identical with $\boldsymbol{\phi}$.

(Recall that the boldface refers to a relational structure). Put in this way, the inference is just fine. If the two domains are isomorphic, they instantiate exactly the same mathematical, relational structure, hence, they are qualitatively identical in this respect. However, (I-M)-1 does *not* faithfully capture Köhler's and Petitot's thought, for what they refer to, when they talk about monism, is not a relation of qualitative identity with respect to relational structures, but of *numerical identity* between mind and brain, i.e. between what instantiate such structures. Hence, Köhler's and Petitot's idea might be better captured by:

> (I-M)-2: If $\boldsymbol{\Psi} \cong \boldsymbol{\phi}$, then Ψ is numerically identical with ϕ.

(Recall that the italics refer to carrier sets). Obviously, the consequent of (I-M)-2 naturally entails the following:

> If Ψ is numerically identical with ϕ, then $\boldsymbol{\Psi}$ is numerically identical with $\boldsymbol{\phi}$.

(That is, since the carrier sets are numerically identical, their relational structures must be numerically identical as well. This naturally follows from the application of Leibniz's law). (I-M)-2 is very different from (I-M)-1. The key difference is that in (I-M)-2 there is a jump from an antecedent, which expresses a mathematical function, to a consequent, which expresses a relation of numerical identity between carrier sets. There are two problems with (I-M)-2.

Firstly, the fact that there is an isomorphic function between the relevant domains does *not* justify the inference to numerical identity of the sets. Indeed, there are many examples of different domains or objects, mathematically described, which are numerically different. Consider again the case of the set of natural numbers and the sequence of annual time segments (§3.1). Of course, no one would say that the sequence of annual time segments is numerically identical with the sequence of natural numbers, even if they instantiate the same relational structure. Putting things more informally, we can say that from the fact that two things instantiate the same property (e.g. "being blue") it obviously does not follow that they are numerically identical (e.g. your shirt and the sky). One can reply that my interpretation is uncharitable, perhaps, neither Köhler nor Petitot think that (I-M)-2 brings *conclusive* evidence for monism. Rather, their claims should be interpreted as saying that, *if* we could show that $\mathbf{\Psi}$ is not isomorphic with ϕ, they could not be numerically identical, again, in compliance with Leibniz's law. However, as we are about to see (§4.3), things are further compounded by multiple possible ways to mathematically describe the relevant structures.

Secondly, further reflection suggests that even (I-M)-2 does not faithfully capture Köhler's and Petitot's ideas. Let us zoom in on the consequent. The consequent expresses a relation of numerical identity between carrier sets. But carrier sets just cannot be the "neural" and the "mental," for sets are abstract mathematical entities! The carrier sets of the respective relational structures are just mathematical constructs, or, if we want, sets of symbols used to refer to worldly things in the world[12]. To make this point clear, consider the following example. Suppose you want to draw a list of all the people who sit in your living room right now. (Such a list might, of course, be empty). The list contains all and only the names of people in your living room, but the list contains obviously just names. The list might also be ordered, for example we may put the names in alphabetical order. However, what you would sort in this case are names, not real people in your living room. The most obvious implication of this problem is that talk about isomorphism is confined within mathematical entities, whereas talk about the alleged mind-brain identity refers to things in the world. I will further elaborate the consequence of this insight in §4.3.

4.2. From Monism to Isomorphism

A clear expression of this strategy can be found in Revonsuo:

> …there must be isomorphism between one specific level of organization in the brain and phenomenal consciousness, simply because these boil down to one and the same thing. (2000, p. 67).

---

[12] A further complication here is to determine which symbols stand in for worldly entities and which ones are merely internal to the representational system, but we can skip this complication here.

Once more, there is no further specification about the kind of identity assumed. Furthermore, Revonsuo does not discuss in what sense phenomenal consciousness would be structured, and thus fails to meet the second requirement of isomorphism (§3.1). We can abstract away from these issues, and zoom in, again, on the structure of this claim. Adopting our familiar terminology, the claim can be regimented as follows:

> (M-I): If the mind is the brain (*Monism*), hence Ψ is numerically identical with ϕ, then $\boldsymbol{\Psi}$ *must* be isomorphic with $\boldsymbol{\phi}$ (Automorphism).

Clearly, the implication allows for the consequent to be true even in the falsity of the antecedent: two numerically distinct things can be isomorphic. What is interesting is whether the consequent must follow from the antecedent. Now, as we have seen (§3.1), the fact that domain and image are numerically identical does not per se warrant that just any function h will be invertible, and thus an automorphism, for it is thoroughly possible that an h from $\mathbf{A}$ to $\mathbf{A}$ (or from $\boldsymbol{\Psi}$ to $\boldsymbol{\phi}$) will be a mere *endomorphism*. This was precisely the point illustrated by means of our example of endomorphism from a vector space $\mathbf{V}$ to $\mathbf{V}$. A cheap response may be that if domain and image are numerically identical, then it will always be possible to specify an identity function, i.e. a function whose output just corresponds to the input value. In such a case, however, psychoneural isomorphism will not be an interesting thesis, all it would give us is simply the value we already know! Beside the identity function, however, the exact function at stake must be further specified in order to see whether it is an automorphism or not. In other words, it is not obvious that given the numerical identity of the domains an automorphism follows. This may seem odd at first, but careful reflection suggests that the source of our puzzlement comes from mistaking the third requirement of isomorphism for a metaphysical intrinsic relation. An isomorphism, like any other morphism, is a mathematical function, it is a process we use to get an output once we fix a value chosen from the domain and as such it operates between abstract mathematical models, not things in the world.

4.3. Mathematical Models and Their Roles

The foregoing considerations put pressure against the ontic reading of isomorphism. Carrier sets, relational structures, and functions are abstract mathematical concepts. So, how can we make sense of psychoneural isomorphism in the first place? The short answer is, via mathematical models. Let our worldly things be the model's target. A mathematical model is an interpreted, idealized mathematical structure that stands in some representational relation to its targets and that can be studied to gain indirect insights about the targets they are about (Frigg & Hartmann 2009; Giere 1988; Weisberg 2013). It is only between such mathematical structures that we may find an isomorphism.

How should we model the targets? There are no strict rules for doing so. In general, mathematical models, just as other scientific models, are not meant to be mirror images of their targets. Models contain idealizations, abstractions (Weisberg 2007). Wisely contrived, such distortions enhance the epistemic power of our models (Elgin 2017, pp. 23-32). The way we build a mathematical model, just like any other model, and therefore what to leave out and what parameters should be idealized, depends on our *epistemic goals*. Usually, a model devised to maximize an epistemic goal does so at the expenses of other goals. Some models may have purely explorative value (e.g. Gelfert 2016 and below), others may maximize descriptive accuracy while having little predictive power, whereas other models provide scientific explanations. Scientific models play many other roles as well, but we will just focus on a basic distinction that is later going to play an important role (§§5.2-3). Some models play *explanatory* roles, others do not. Models of the latter kind are often called "phenomenological" (Frigg & Hartmann 2009; Hochstein 2013; Wimsatt 2007), but in order to avoid confusions with other uses of the term "phenomenology," I shall simply call them *non-explanatory* models.

Non-explanatory models have different uses. Batterman for instance examined the role of minimal models—i.e. highly idealized models—in statistical mechanics, and concludes that the best way to think of their role is «that they are means for extracting stable phenomenologies [i.e. regularities] from unknown and, perhaps, unknowable detailed theories» (2002, p. 35). Such regularities may then be used for computational or explanatory purposes (ivi, p. 37). Another fitting example comes from Bogen (2005). Following Mitchell's contention that the role of scientific generalizations «is to provide reliable expectations of the occurrence of events and patterns of properties» (2003, p. 124), Bogen argues that such models may be used to:

- Describe facts to be explained;
- Suggest constraints on acceptable explanations;
- Suggest and sharpen questions about causal mechanisms;
- Measure or calculate quantities;
- Support inductive inferences (2005, p. 401).

As an example, he considers the famous Hodgkin-Huxley equations of action potential, the pulse of electricity that traveling down the axon towards the synapse triggers the release of neurotransmitters. Studying the squid giant axon, Hodgkin and Huxley argued that the magnitude of the potassium current I_K which help repolarize the membrane varies with $\bar{g}_K$, the membrane's maximum potassium current conductance, a weighting factor (n^4), and a driving electrical force equal to the difference between E_m, the membrane potential, and the resting potential for potassium, E_k. The equation for the potassium is:

$$I_K = n^4 \, \bar{g}_K \, (E_m - E_K)$$

As Bogen argues, this equation incorporates the «qualitatively correct idea that I_K varies with $(E_m - E_K)$», yet he specifies that this model is also «quantitatively inaccurate to a significant degree». But in spite of its inaccuracies and poor *predictive* power, the model has played an important role for studying action potential. The mechanism governing action potential was, at that time, still unknown, and Hodgkin and Huxley meant their equations to be «empirical descriptions» of the target phenomenon (Hodgkin & Huxley 1952, p. 541; quoted from Bogen 2005, p. 404). The model served a useful *exploratory* role, describing the behavior of the phenomenon and, as Bogen says, indicated the «features of the phenomena of interest which mechanistic explanations should account for» (ivi, p. 403; cf. also Gelfert 2016, pp. 79-97).

The fact that different models embody different epistemic purposes directly bears on the case of psychoneural isomorphism, for when we construct a mathematical model our epistemic goals will determine which mathematical structure will be relevant, and accordingly, whether two models will be isomorphic or not. We can illustrate this point by means of a popular example. Suppose we take your coffee mug on the desk and that donut you bought for breakfast. How should the mathematical models capture the targets' structures? This depends on our epistemic goals. Within a classical geometrical model, clearly, the mug and the donut (a torus) do not have the same structure, hence our models will carry relational structures which clearly are not isomorphic. However, if we are interested in topological spaces (Munkres 2000, p. 76), things will be very different. It is one of the most quoted examples in topology that a coffee mug is homeomorphic (i.e. topologically isomorphic, §3.1) to a torus (the donut). Of course, it depends on our epistemic goals which mathematical model we will have to construct: exploratory, descriptive, explanatory, etc. (We will further explore these considerations in §5), and in turn this will determine whether our models will be isomorphic or not.

Time to take stock, the correct analysis of psychoneural isomorphism must be an epistemic reading where mathematical models should be sharply distinguished from their targets. Let our targets be, as we have seen, S's conscious visual content $\mathfrak{C}$ at t, and the underlying neural structure $\mathfrak{N}$ that sustains it at the same time. A mathematical model E of $\mathfrak{C}$ will specify the carrier set Ψ together with a relational structure $\boldsymbol{\Psi}$ for epistemic purpose P; a mathematical model M of $\mathfrak{N}$ will specify a carrier set ϕ together with a relational structure $\boldsymbol{\phi}$ for an epistemic purpose P'. In the next section, I will focus on the alleged roles of isomorphism within the epistemic reading, focusing on Petitot's morphodynamical approach.

5. From Morphodynamics to Robustness

In this section, I am mostly going to focus on the work of the French mathematician and philosophers Jean Petitot. The are two reasons for this. Firstly, because he has provided the single most developed *mathematical* account of psychoneural isomorphism. Second, because in such an account Petiot has embraced all putative roles of psychoneural isomorphism, identifying his contribution as both in the project of naturalized Phenomenology (§2.2.1) and the search for the neural correlates of conscious content (§2.2.2). Thus, his work provides an ideal case-study for my purposes. Setting the ontic reading aside, we will now look closer (§5.1) at Petitot's approach and examine the alleged epistemic roles of isomorphism (§§5.2-3).

5.1. Petitot's Neurogeometry

Petitot is mainly interested in specifying the neurogeometry of the functional architecture of visual areas; more precisely the problem is the:

> …implémentation neuronale des algorithms de cette géométrie, le problem étant de comprendre comment les structures perceptives "macro" et leur morphodynamique peuvent émerger du niveau "micro" neuronal sous-jacent (2008, p. 22).
> [the neural implementation of this geometry's algorithms, the problem being that of understanding how the "macro" perceptive structures and their morphodynamics can emerge from the underlying "micro" neural level.]

This proposition is inscribed within the larger project of naturalizing Phenomenology (§2.2.1) in which Petitot occupies a unique position, having developed the best-articulated mathematical account of the program. According to the programmatic statements in Roy et al. (1999), the ontological divide between the "mental" and the "physical" can be bypassed by means of a mathematization of the two[13]. The mathematization, so understood, represents a way of naturalizing Phenomenology, i.e. a way of making Phenomenology continuous with the natural sciences. The mathematization follows these steps:

1) Phenomenological descriptions.
2) Mathematical models.
3) Naturalistic model.

The passage from (1) to (2) is achieved via a "theory-theory" or "model-model" «exact correspondence» (Petitot 1999, p. 343; §5.2). Phenomenological descriptions are conceptual, whereas the corresponding geometrical eidetics is expressed by means of a morphodynamical model (2008, p. 395). We thus obtain the following schema (Tab. 2).

Conceptual Eidetics	**Geometrical Eidetics**
Phenomenological Descriptions	Morphodynamical Models

Tab. 2: Conceptual and geometrical eidetics.

Let us consider an example. Petitot focuses on Husserl's insightful discussion of phenomenal saliency (*phänomenale Abhebung*), which enables the individuation of phenomena, i.e. appearances (cf. Husserl 1993, pp. 242-245). Phenomenal saliency is made possible by means of a distinction between distinct (*gesondert*) contents and fused (*verschmolzen*) contents. The process of fusion (*Verschmelzung*) creates a phenomenal whole; whereas the opposite process of distinction (*Sonderung*) demarcates the different parts. The *Sonderung* is based on the qualitative discontinuity of the "moments"—roughly, particularized properties—that compose the contents[14]. In short, the structure of visual appearances is

---

[13] It is a separate and interesting question to assess whether Phenomenology may be mathematized (e.g. Zahavi 2004).

[14] Mulligan (1999) has argued that Husserl's moments are trope-like entities.

based on the qualitative discontinuities between moments. These concepts, however, are not continuous with the concepts employed by neuroscientists in their researches, this is precisely why we need to bridge this conceptual gap.

Petitot argues that a mathematical translation of Husserl's analysis is possible (step 1 to 2). The relation between quality (e.g. color) and space corresponds to a (mathematical) category (Petitot 1999, p. 339; cf. also 1993; 2011, pp. 64-65), and in particular to a fibration or fibred space. A fibration is a differentiable manifold E endowed with a canonical projection $\pi: E \rightarrow M$ (a differentiable map) over another manifold M. $E_x = \pi^{-1}(x)$of the points $x \in W$ by π are the "fibres" of the fibration, subspaces of E that are projected to points in M. A fibration must satisfy two axioms:

1. All the fibres E_x are diffeomorphic with a typical fiber F.
2. The projection π is locally trivial, i.e. for every $x \in M$, there exists a neighborhood U of x such that the inverse image $E_U = \pi^{-1}(U)$ of U is diffeomorphic with the direct product $U \times F$ endowed with the canonical projection $U \times F \rightarrow U, (x, q) \rightarrow x$.

(A diffeomorphism is an isomorphism between manifolds, a topological space). This mathematical model would capture the relation between quality and extension in Husserl's Phenomenology (Husserl 1991, pp. 68-71; Petitot 2004). The base of the fibration is the extension and the total space is a sensible quality, say, color. With this mathematical model of $\mathfrak{C}$ (perceptual content, or better, an aspect of it) inspired by Phenomenological concepts, we have specified a carrier set and relational structure. We now need to move from step 2 to step 3. More precisely, we need to pin down some physical-mathematical model of the neural dynamics that implements the geometric description (Petitot 1999, pp. 338-343; 2008, pp. 380-381).

Petitot argues that one of the main problems of natural and computer vision is to understand «how signals can be transformed into geometrically well-behaved observables» (1999, p. 346), i.e. the process whereby an unstructured image I(x, y) becomes segmented. Perhaps the most widespread mathematical model for segmenting an image into distinct parts has been developed by Mumford (1994), and it is known as the Mumford-Shah model. There are alternative models as well, more local and based on anisotropic non-linear partial differential equations (Petitot 1999, p. 348; 2011, pp. 78ff). (I skip the mathematical details, the reader interested can find them in Petitot 1999, 2008, 2011, 2013). The relevant point is that the same fibration used to model the Phenomenological descriptions can be used to model the neurogeometry of the functional architecture of V1, the primary visual cortex. More specifically, Petitot develops his account basing on Hubert & Wiesel's (Bechtel 2001, pp. 232-234) discovery of the micromodules called hypercolumns (Petitot 2008, 2013, p. 75)[15]. We have thus achieved a genuine *psychoneural isomorphism* that respects all three requirements (§3.1):

> […] l'accord entre le macro-niveau géométrique (morphologique) émergent M [...] et l'expérience phénoménale E [...] est extrêmement fort, beaucoup plus fort qu'une simple corrélation. C'est même la forme la plus forte possible de matching de contenus puisque, à la limite, c'est un isomorphisme» (2008, p. 367; emphasis added).
> [The matching between the emergent geometrical macro-level M (morphology) […] and the phenomenal experience E […] is very strong, much stronger than a simple correlation. It is the stronger possible kind of content matching since, at its limit, it is an isomorphism.]

What is, however, the epistemic achievement of such an isomorphism?

Firstly, Petitot contends that «[w]ith such a morphodynamical model we can easily *explain* the topological description physically» (2011, p. 69; emphasis added). The model, apparently, extends its explanatory virtue also to the problem of subjective contours (the Kanizsa triangle, for example) or phenomena like the neon color spreading (§2.2.2.), the subjective impression of a color spreading across the four circles represented in the Neon Color Spreading (cf. Petitot 2003, 2013, pp. 81ff). In short, such models would explain «the structure of percepts» (Petitot 2013, p. 75). A mathematical (topological) or, as I shall say, following Haugeland (1998), *morphological* explanation ensues in virtue of the isomorphism. Morphological explanations are explanations «where the distinguishing marks of the style are that an ability is explained through appeal to a specified structure and to specified abilities of whatever is so structured» (ivi, p. 12). This corresponds to the *Explanatory Role* (§2.3. Tab. 1).

Secondly, Petitot contends that the mathematized Phenomenological descriptions enable us to bridge the conceptual gap between disciplines, using the first-person descriptions as constraints on the admissible naturalistic explanations and models (1999, p. 330). This contention exemplifies two further roles of isomorphism. The first is its *intertheoretic role*, i.e. the problem of showing how different disciplines interact (say, psychology and neuroscience). The second is the *heuristic* role since with the aid of mathematical models of first-person contents, it is claimed that we can guide the search for the underlying neural structures.

5.2. The Epistemic Roles of Psychoneural Isomorphism

I lump together the Intertheoretic and the heuristic role in §5.2.1; I turn then to the explanatory role in §5.2.2.

5.2.1 Intertheoretic and Heuristic Role. Coherently with the project of naturalizing Phenomenology, Petitot conceives phenomenological descriptions as playing a heuristic role in the search for content-NCCs. This would be possible in virtue of

---

[15] Petitot points out that Noë & Thompson (2004)'s negative assessment of psychoneural isomorphism is largely based on the mistaken assumption that single cells would be the neural correlates of perceptual content. Petitot's neurogeometry is based instead on a morphodynamical analysis of larger population of neurons.

the isomorphism between models of neural activity and models of phenomenological descriptions. The relation between different models or theories—sometimes between different levels of description[16]—is known as the problem of *intertheoretic relations* (e.g. Danks 2014)[17]. The classical account of intertheoretic relation is *intertheoretic reduction* (e.g. Schaffner 1993).

Intertheoretic reduction follows the schema of deductive-nomological explanations (DN). On this account, scientific explanations are deductive arguments in which the explanandum features as the conclusion and among the premises there must be at least one law of nature (Hempel & Oppenheim 1948; Salmon 1989). An application of this model to intertheoretic reduction in neuroscience and its relation to DN explanations can be found in Churchland (1986). The reduction at stake here is a relation that obtains between models (or theories). Model reduction is sometimes thought to lead to ontological reduction of the models' targets. In order to clarify Churchland's thesis, and its relation with our main point, let us examine a classical example. Nagel (1961, pp. 338-345) distinguished between two types of reduction: homogeneous and heterogeneous. In the former case, the "primary" theory (reducing) and the "secondary" theory (reduced) share the same vocabulary. This allows a simple reduction where the reduced theory features as conclusion of a deductive argument. Things are different in the case of heterogeneous reduction. The classical example is the reduction of thermodynamics to statistical mechanics (ivi, pp. 339-345), where the former contains terms and concepts such as "temperature" that are absent in the primary theory. Without terminological consistency, it is not possible to establish a logical-deductive relation. In order to overcome this obstacle, Nagel (ivi, 353-354) introduced a «condition of connectability» that bridges the gaps between primary and secondary theory. Back to Churchland. The reduction of the mental vocabulary to a neural vocabulary poses an obvious challenge of heterogeneity. In order to bridge the terminological and conceptual gap of the mental vocabulary within theory T_F, however, we can create a an isomorphic model of T_F, call it T_F*, which in turn can be deduced from a theory of neural processes T_N following the schema:

$$T_N \rightarrow T_F^* \cong T_F$$

(The arrow here does not represent the logical connective of material implication, but a deducibility relation). My contention is that Petitot's approach is strikingly similar to Churchland's. In Petitot's case a phenomenological, conceptual description D of $\mathfrak{C}$ (perceptual content), serves as base for creating a mathematical model E, which specifies a carrier set $\mathit{\Psi}$ and a relational structure $\boldsymbol{\Psi}$. This is

---

[16] Some reductions are intra-level, as in the case of theories or models within the same level of description; our focus here is on models that belong to two different levels of description, i.e. the experienced or phenomenological, and the neural (Nickels 1973).

[17] Much of the philosophical literature has focused on relations between theories, but the same considerations apply to models as well.

roughly an equivalent to the right hand side of Churchland's schema. In addition, from $\mathfrak{N}$ we get a mathematical model M that specifies a carrier set ϕ together with a relational structure $\boldsymbol{\phi}$. We thus obtain an isomorphism $\boldsymbol{\phi} \cong \boldsymbol{\Psi}$ of the respective neural and phenomenological models.

Can this account for the *heuristic* role of phenomenological descriptions, as Köhler and proponents of naturalized Phenomenology urge? The short answer is no, and there are two main reasons for this. First, because this isomorphism has a *reconstructive* character. The relevant neural structures underlying conscious perception must have been previously singled out in order for us to build a mathematical model thereof. The discovery of such structures does not rely on isomorphism or mathematical models, but is mostly achieved via careful selective interventions (Craver 2007; Woodward 2003) that uncover the constitutive or causally relevant components of the target system that bring about a change in the phenomenon, i.e. in our case conscious perceptual content. Second, because the isomorphism holds only between very specific mathematical models, and not between any mathematical model of the neural structures' activities or of perceptual content. And the choice of models, as we have seen (§4.3), largely depends on our epistemic goals. In general, and most of the time, cognitive scientists rely on a plurality of different models that serve different epistemic purposes and that target different facets of the phenomenon.

It may be argued that the approach vindicates the intertheoretic role of isomorphism. After all, as Petitot, Varela, Roy et al., and Köhler insisted (§2.1.2; §2.2.1), phenomenological descriptions are meant to deepen our understanding of how the brain works from a rigorous first-person perspective by putting constraints on models of the neural. Yet, there is a tension between Petitot's approach and the claim that phenomenological descriptions should put constraints on neural models. It lies in the fact that intertheoretic constraints are usually conceived as a better alternative to intertheoric reduction (e.g. Craver 2005, 2007, pp. 256ff; Danks 2014). While a lengthy discussion of this issue must be postponed to another contribution for reasons of space (cf. Vernazzani 2016), the following observation by Craver nicely summarizes the core issue: neuroscientists do not «create a homomorphic image of a phenomenon studied by those in another field» (2007, p. 266). The price of intertheoretic reduction is abstracting away from current neuroscience practice to achieve some sort of normative ideal, one that, perhaps, better suits more abstract epistemic purposes, like the quest for the unity of science (Oppenheim & Putnam 1958). As a regulative ideal, however, such intertheoretic reduction flies in the face of more local approaches.

5.2.2. A Morphological Explanation? Petitot maintains that his account provides an explanation of perceptual content's structure. In his 2008, he even claims that the isomorphism between **M** and **E** bridges the explanatory gap (Levine 1983), showing why perceptual content has the consciously "felt" character.

The structure of the specific model of scientific explanation, however, is far from clear. Petitot seems to suggest that explanation has to do with deduction or derivation (Petitot 2008, p. 31). This draw his account close to what Haugeland called the «derivational-nomological» style of explanation, i.e. a «special case form of deductive-nomological explanation—where the distinction of the special case is that the presupposed regularities are expressed as equational relationships among quantitative variables, and the deduction is mathematical derivation of other such equations» (1998, p. 11; cf. §5.2.1). However, he also characterizes his approach as a kind of mathematical explanation or, to use Haugeland's terminology again, as a «morphological explanation» whose hallmark is that «an ability is explained through appeal to a specific structure and to specified abilities of whatever is so structured» (1998, p. 12). This is particularly clear in his characterization of the fibration structure of perceptual content as deriving mathematically from the structure of the underlying neurogeometry of V1 (Petitot 2013, 2008).

A first problem with Petitot's explanatory ambitions consists in the ambiguous nature of the explanandum. Put more informally, is the morphodynamical account supposed to provide an answer to the question "Why does the brain produce that particular perceptual content's structure?" or is it "Why do hypercolumns's neural activity mathematically necessitates that particular perceptual structure?" or, again, "Why is that particular perceptual content's structure conscious?" These are only a few possible questions; the point is that different ways of singling out the explanandum will require different explanations. Furthermore, whether isomorphism provides an explanation or not will also depend on the norms of explanation accepted within the given scientific community relative to a given account of explanation. For instance, in the case of neuroscience and cognitive science, several researchers (e.g. Bechtel & Richardson 2010; Craver 2007; Craver & Darden 2013; Miłkowski 2013) have convincingly argued that explanation is often (if not always) construed as a search for mechanisms, i.e. structured entities whose activity constitutes or causes the explanandum phenomenon. On this perspective, a mechanistic explanation of *x* will consist in determining all (and only) the constitutively or causally relevant components parts and operations of the responsible mechanism. Such explanations admit degrees of completeness, ranging from merely how-possibly sketches to how-actually, complete descriptions for the specific explanatory purpose (Craver & Darden 2013).

Morphological explanations' relation with mechanistic accounts of explanation has recently drawn some attention (e.g. Huneman 2018; Lange 2013; Levy & Bechtel 2013; Rathkopf 2015). My contention is that the explanatory role of isomorphism is unclear. Consider again our example of the mug and the donut: they are homeomorphic, but so what? No one would conclude that one thing explains the other or its structure. In our case, Petitot's model of the hypercolumns' activity may be a mathematically impressive achievement, but it does not help *explain* how the relevant neural structures' computations achieve this, not even if one assumes that they are numerically identical entities. In order to do so, one would have to clarify the nature of the relevant neural parts and operations and their organization. This brings us to a second problem.

As we have seen (§3.3), models are devised to serve some epistemic goal. In several passages, Petitot (2008) contends that his mathematical translation of phenomenological descriptions (steps 1 to 2) conform to a rigorous specification of Marr's *computational* level. Marr described the scientists' task at the computational level as «an abstract formulation of what is being computed and why» (1977, p. 37; cf. also Marr 2010). The exact interpretation of Marr's computational level has been object of intense debate among philosophers and cognitive scientists (e.g. Shagrir & Bechtel 2017). According to some interpreters, the computational level consists solely in the specification of the task to be solved by the information-processing system (ivi, pp. 193-194). Egan (1991, 1995) has put forward an interpretation of Marr's computational level that consists in the mathematical specification of the function(s) computed (1995, p. 185). Petitot's own approach can be understood along the lines of Egan's interpretation, as his model provides a rigorous mathematical characterization of the problem to be solved, i.e. understanding how the neural system carries out the targeted function (Petitot 2008, p. 22). Put in these terms, however, and without disputing Marr's exegesis, the morphodynamical model E of perceptual content (its target) is a "non-explanatory" model that provides a mostly accurate *mathematical description* of the target.

Let us now turn to the mathematical model of the underlying neural activity, M. Such a model, once more, does not specify how the target neural structure actually performs the computations, but provides a mostly accurate mathematical model of the neural structure's activity as a whole. The isomorphism between M's and E's relational structures, in short, does not seem to embody any explanatory epistemic goal. This is not to say that they M and E are theoretically idle, they might play a variety of different non-explanatory roles. It is to one among such roles that I now turn.

5.3. Psychoneural Isomorphism and Robustness

In every explanation, including of course mechanistic explanation, the characterization of the explanandum phenomenon plays a central role (e.g. Bechtel 2013; Shagrir & Bechtel 2017). How to correctly characterize or describe conscious phenomena is anything but simple. One option is to construe $\boldsymbol{\phi} \cong \boldsymbol{\Psi}$ as a step in the phenomenal stabilization or robustness of the explanandum (e.g. Feest 2011; Wimsatt 2007). The notion of stabilization refers to:

- The processes and methods whereby scientists empirically identify a given phenomenon, and
- Gradually come to agree that the phenomenon is a stable, and robust feature of the world, rather than an artifact produced by an instrument, methodological assumptions and procedures, etc.

One (or perhaps, the *only*) way to determine the robustness of a given phenomenon is by means of multiple determinations in its identification, i.e. the convergent results between different methodologies or levels of analysis (cf. Hacking

1983; Wimsatt 2007, pp. 37-74). The need for robustness regarding mental perceptual phenomena, such as filling-in (§2.2.2) is motivated by the unreliability of first-person, naïve descriptions of one's own perceptual experiences (Dennett 1991; Schwitzgebel 2011). Notice that this was the very same motivation that grounds the introduction of rigorous first-person methodologies at the base of the project of naturalized Phenomenology (§2.2.1). Within this context the isomorphism $\boldsymbol{\phi} \cong \boldsymbol{\Psi}$—if achieved through independent modeling of each of the two targets—may provide evidence for the robustness of the phenomenon under scrutiny. It may do so by dint of the independent achievement of the same descriptions of the target phenomenon relying on different sources of data. Surely, the isomorphism may not be (and we should not expect it to be) the only method for achieving phenomenal stabilization, but it may represent a new and helpful conceptual tool in stabilizing the target phenomenon. This further role deserves to be further examined in subsequent studies.

6. Conclusion

In this contribution, I set out to shed light on the concept of psychoneural isomorphism. I have provided a taxonomy of the putative roles of isomorphism and dispelled some misunderstandings regarding the relation between isomorphism and the metaphysics of the mind. I have urged that psychoneural isomorphism is better understood as a function between mathematical models, and stressed the importance of multiple, not always compatible, epistemic roles that such models embody. I have later reviewed the epistemic roles of psychoneural isomorphism, using the work of Jean Petitot as a case-study. While my conclusion has been largely negative, I hinted at a possible role of psychoneural isomorphism in phenomenal stabilization of perceptual content that should be object of further studies.

References

Allen, S. (2016). *A Critical Introduction to Properties*. London-New York: Bloomsbury.

Batterman, R. (2002). "Asymptotics and the Role of Minimal Models" *British Journal for the Philosophy of Science* 53, pp. 21-38.

Bechtel, W. (2001). "Decomposing and Localizing Vision: An Exemplar for Cognitive Neuroscience."In W. Bechtel, P. Mandik, J. Mundale & R. Stufflebeam (eds.) *Philosophy and the Neurosciences* (pp. 225-249). Malden: Blackwell.

— (2013). "Understanding Biological Mechanisms: Using Illustrations from Circadian Rhythm Research." In K. Kampourakis (ed.) *The Philosophy of Biology* (pp. 487-510). Dordrecht: Springer.

Bechtel, W. & R. Richardson (2010). Discovering Complexity. Cambridge, MA: MIT Press.

Bogen, J. (2005) "Regularity and Causality: Generalizations and Causal Explanation." *Studies in the History and Philosophy of Biology and Biomedical Science* 36, pp. 397-420.

Bressan, P. Mingolla, E., Spillmann, L. & Watanabe, T. (1997). "Neon Color Spreading: A Review." *Perception* 26, pp. 1353-1366.

Brewer, B. (2011). *Perception and Its Objects*. Oxford: Oxford University Press.

Bridgeman, B. (1983). "Isomorphism is where you find it." *Behavioral and Brain Sciences* 6, pp. 658-659.

Byrne, A. (2001). "Intentionalism Defended" *The Philosophical Review* 110(2), pp. 199-240.

Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.

— (2000). “What Is a Neural Correlate of Consciousness?” In T. Metzinger (ed.) *Neural Correlates of Consciousness* (pp. 17-39). Cambridge, MA: MIT Press.

— (2012). “A Computational Foundation for the Study of Cognition” *Journal of Cognitive Science* 12, pp. 323-357.

Churchland, P.S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.

Cohn, P.M. (1981). *Universal Algebra*. Dordrecht: Reidel.

Craver, C. (2005). “Beyond Reduction: Mechanisms, Multifield Integration, and the Unity of Neuroscience.” *Studies in History and Philosophy of Biology and Biomedical Sciences* 36, pp. 373-395.

— (2007) *Explaining the Brain*. New York: Oxford University Press.

— (2014). “The Ontic Account of Scientific Explanation.” In M.I. Kaiser, O.R. Scholz, D. Plenge & A. Hüttermann (eds.) *Explanation in the Special Sciences* (pp. 27-52). Springer Verlag.

Craver, C. & L. Darden (2013). *In Search of Mechanisms*. Chicago: University of Chicago Press.

Crick, F. & C. Koch (1995). “Are We Aware of Neural Activity in Primary Visual Cortex?” *Nature* 375, pp. 121-123.

— (1998). “Consciousness and Neuroscience.” *Cerebral Cortex* 8, pp. 97-107.

Danks, D. (2014). *Unifying the Mind*. Cambridge, MA: MIT Press.

Darden, L. & N. Maull (1977). “Interfield Theories.” *Philosophy of Science* 44(1), pp. 43-64.

Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, & co.

Dunn, M.J. & G. Hardegree (2001). *Algebraic Methods in Philosophical Logic*. Oxford: Clarendon Press.

Egan, F. (1991). “Must Psychology be Individualistic?” *Philosophical Review* 100, pp. 179-203.

— (1995) “Computation and Content.” *Philosophical Review* 104, pp. 181-203.

Elgin, C. (2017). *True Enough*. Cambridge, MA: MIT Press.

Feest, U. (2011). “What exactly is stabilized when phenomena are stabilized?” *Synthese* 182, pp. 57-71.

Flanagan, O. (1992). *Consciousness Reconsidered*. Cambridge, MA: MIT Press.

Frigg, R. & S. Hartmann (2009). “Models in Science.” In E. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/archives/spr2017/entries/modelsscience/

Fry, G.A. (1948). “Mechanisms Subserving Simultaneous Brightness Contrast.” *American Journal of Optometry and Archives of American Academy of Optometry* 25(4), pp. 162-178.

Gelfert, A. (2016). *How To Do Science with Models*. Dordrecht: Springer.

Giere, R. (1988). *Explaining Science*. London: University of Chicago Press.

Greenwood, J. (2015). *A Conceptual History of Psychology*. Second Edition. Cambridge: Cambridge University Press.

Hacking, I. (1983). *Representing and Intervening*. Cambridge, MA: Cambridge University Press.

Haugeland, J. (1998). *Having Thought*. Cambridge, MA: Harvard University Press.

Heidelberger, M. (2000). “Fechner und Mach zum Leib-Seele-Problem.” In A. Arndt & W. Jaeschke (eds.) *Materialismus und Spiritualismus: Philosophie und Wissenschaft nach 1845* (pp. 53-67). Hamburg: Meiner.

Hempel, C.G. & P. Oppenheim (1948). (1948). “Studies in the Logic of Explanation.” *Philosophy of Science* 15(2), pp. 135-175.

Hochstein, E. (2013). “Intentional Models as Essential Scientific Tools.” *International Studies in the Philosophy of Science* 27(2), pp. 199-217.

Huneman, P. (2018). “Diversifying the Picture of Explanations in Biological Sciences: Ways of Combining Topology with Mechanisms.” *Synthese*. https://doi.org/10.1007/s11229-015-0808-z.

Husserl, E. (1991). *Ding und Raum*. Hamburg: Felix Meiner Verlag.

— (1993). *Logische Untersuchungen*, Vol. II/1. Tübingen: Max Niemeyer Verlag.

Köhler, W. (1929). *Gestalt Psychology*. Oxford: Liverlight.

Komatsu, H. (2006). “The Neural Mechanisms of Perceptual Filling-in.” *Nature Reviews Neuroscience* 7, pp. 220-231.

Kriegeskorte, N. & R.A. Kievit (2013). “Representational geometry: integrating cognition, computation, and the brain.” *Trends in Cognitive Sciences* 17(8), pp. 401-412.

Kriegeskorte, N., Marieke, M. & P. Bandettini (2008). “Representational similarity analysis – connecting the branches of systems neurosciences” *Frontiers in Systems Neurosciences*. doi: 10.3389/neuro.06.004.2008.

Lange, M. (2013). “What Makes a Scientific Explanation Distinctively Mathematical?” *British Journal for the Philosophy of Science* 64, pp. 485-511.

Lehar, S. (1999). “Gestalt Isomorphism and the Quantification of Spatial Perception.” *Gestalt Theory* 21(2), pp. 122-139.

— (2003). “Gestalt Isomorphism and the Primacy of the Subjective Conscious Experience: A Gestalt Bubble Model.” *Behavioral and Brain Sciences* 26, pp. 375-444.

Levine, J. (1983). “Materialism and Qualia: The Explanatory Gap.” *Pacific Philosophical Quarterly* 64, pp. 354-361.

Levy, A. & W. Bechtel (2013). “Abstraction and the Organization of Mechanisms.” *Philosophy of Science* 80, pp. 241-261.

Luccio, R. (2010). “Anent Isomorphism and Its Ambiguities: From Wertheimer to Köhler and Back to Spinoza.” *Gestalt Theory* 32(3), pp. 208-234.

Mach, E. (1865). “Über die Wirkung der räumlichen Vertheilung des Lichtreizes auf der Netzhaut.” *Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Classe* 52(2), pp. 303-322.

Madden, E.H. (1957). “A Logical Analysis of ‘Psychological Isomorphism’.” *The British Journal for the Philosophy of Science* 8, pp. 177-191.

Marr, D. (1977). “Artificial Intelligence: A Personal View.” *Artificial Intelligence* 9, pp. 37-48.

— (2010). *Vision*. Cambridge, MA: MIT Press.

Miłkowski, M. (2013). *Explaining the Computational Mind*. Cambridge, MA: MIT Press.

Mitchell, S. (2003). *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.

Müller, G. (1896). “Zur Psychophysik der Gesichtsempfindungen. Kap. 1.” *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 10, pp. 1-82.

Mulligan, K. (1999). “Perception, Particulars and Predicates.” In D. Fisette (ed.) *Consciousness and Intentionality* (pp. 163-194). Dordrecht: Kluwer.

Mumford, D. (1994). “Bayesian Rationale for the Variational Formulation.” In B. M. ter Haar Romney (ed.) *Geometry-driven Diffusion in Computer Vision* (pp. 135-146). Dordrecht: Kluwer Verlag.

Munkres, J. (2000). *Topology*. Uppder Saddle River: Prentice Hall.

Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace & World.

Neuroscience.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 373-395.

Nickles, T. (1973). “Two Concepts of Intertheoretic Reduction.” *The Journal of Philosophy* 70(7), pp. 181-201.

Noë, A. & Thompson, E. (2004). “Are There Neural Correlates of Consciousness?” *Journal of Consciousness Studies* 11(1), pp. 3-28.

Noonan, H. & B. Curtis (2014). “Identity.” In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/archives/sum2014/entries/identity/.

O’Regan, K. (1992). “Solving the ‘Real’ Mysteries of Visual Representations: The World as an Outside Memory.” *Canadian Journal of Philosophy* 46(3): pp. 461-488.

Oppenheim, P. & H. Putnam (1958). “Unity of Science as a Working Hypothesis.” *Minnesota Studies in the Philosophy of Science* 2, pp. 3-36.

Palmer, S. (1999). “Color, Consciousness, and the Isomorphism Constraint.” *Behavioral and Brain Sciences* 22, pp. 923-989.

Pessoa, L. & P. De Weerd (eds.) (2003). *Filling-in: From Perceptual Completion to Cortical Reorganization*. Oxford: Oxford University Press.

Pessoa, L., Thompson, E. & A. Noë (1998). "Finding Out About Filling-in: A Guide to Perceptual Completion for Visual Science and the Philosophy of Perception." *Behavioral and Brain Sciences* 21, pp. 723-802.

Petitot, J. (1992-1993). "Phénoménologie naturalisée et morphodynamique: La fonction cognitive du synthétique 'a priori'." *Intellectica* 17, pp. 79-126.

— (1999). "Morphological Eidetics for a Phenomenology of Perception." In Petitot et al. (1999), pp. 330-371.

— (2004). "Géométrie et vision dans 'Ding und Raum' de Husserl." *Intellectica* 2, pp. 139-167.

— (2008). *Neurogéométrie de la vision*. Paris: Les Editions de l'École Polytechnique.

— (2011). *Cognitive Morphodynamics*. Bern: Peter Lang.

— (2013). "Neurogeometry of neural functional architectures." *Chaos, Solitons & Fractals* 50, pp. 75-92.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. New York: Oxford University Press.

Pribram, K.H. (1983). "What Is Iso and What is Morphic in Isomorphism?" *Psychological Research* 46, pp. 329-332.

Rathkopf, C. (2015). "Network Representation and Complex Systems." *Synthese*. Doi 10.1007/s11229-015-0726-0.

Ratliff, F. & Sirovich, L. (1978). "Equivalence Classes of Visual Stimuli" *Vision Research* 18(7), pp. 845-851.

Revonsuo, A. (2000). "Prospects for a Scientific Research Program on Consciousness." In Metzinger (2000), pp. 57-75.

Roy, J.-M., Petitot, J., Pachoud, B. & Varela, F. (1999). "Beyond the Gap: An Introduction to Naturalizing Phenomenology." In J. Petitot, F. Varela, B. Pachoud & J.-M. Roy (eds.) *Naturalizing Phenomenology* (pp. 1-80). Stanford: Stanford University Press.

Salmon, W. (1989). *Four Decades of Scientific Explanation*. Pittsburgh, PA: University of Pittsburgh Press.

Schaffner, K.F. (1993). *Discovery and Explanation in Biology and Medicine*. Chicago: University of Chicago Press.

Scheerer, E. (1994). "Psychoneural Isomorphism: Historical Background and Current Relevance." *Philosophical Psychology* 7(2), pp. 183-210.

Scheutz, M. (2001). "Computational versus Causal Complexity." *Minds and Machines* 11, pp. 543-566.

Schwitzgebel, E. (2011). *Perplexities of Consciousness*. Cambridge, MA: MIT Press.

Shagrir , O. & W. Bechtel 2017). "Marr's Computational Level and Delineating Phenomena." In D. Kaplan (ed.) *Explanation and Integration in Mind and Brain Sciences* (pp. 190-214). New York: Oxford University Press.

Shepard, R. & S. Chipman (1970). "Second-Order Isomorphism of Internal Representations: Shapes of States" *Cognitive Psychology* 1, pp. 1-17.

Siegel, S. (2010). *The Contents of Visual Experience*. New York: Oxford University Press.

Teller, D. (1984). "Linking Propositions." *Vision Research* 24(10), pp. 1233-1246.

Thompson, E. (2007). *Mind in Life*. Cambridge, MA: MIT Press.

Todorovic, D. (1987). "The Craik-O'Brien-Cornsweet Effect: New varities and Their Theoretical Implications." *Perception & Psychophysics* 42(6), pp. 545-650.

Varela, F. (1997). "Neurophenomenology: A Methodological Remedy for the Hard Problem." In J. Shear (ed.) *Explaining Consciousness*. Cambridge, MA: MIT Press.

Vernazzani, A. (2016). "Fenomenologia naturalizzata nello studio dell'esperienza cosciente." *Rivista di filosofia* 107(1), pp. 27-48.

Von der Heydt, R., Friedman, H. & Zhou, H. (2003). "Searching For the Neural Mechanism of Colour Filling-in." In Pessoa & De Weerd (2003), pp. 106-127.

Weil, R. & Rees, G. (2011). “A new taxonomy for perceptual filling-in.” *Brain Research Reviews* 67, pp. 40-55.

Weisberg, M. (2007). “Three Kinds of Idealization.” *The Journal of Philosophy* 104(12), pp. 639-659.

— (2013). *Simulation and Similarity*. Cambridge, MA: MIT Press.

Wimsatt, W. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.

Woodward, J. (2003). *Making Things Happen*. New York: Oxford University Press.

Wright, C. (2012). “Mechanistic Explanation without the Ontic Conception.” *European Journal for Philosophy of Science*. DOI 10.1007/s13194-012-0048-8.

Wu, W. (2018). “The Neuroscience of Consciousness.” In E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/entries/consciousness-neuroscience/.

Zahavi, D. (2004). “Phenomenology and the project of naturalization.” *Phenomenology and the Cognitive Sciences* 3(4), pp. 331-47.