Uwe Meixner · Albert Newen (eds.)

Managing Editor: Pieter Sjoerd Hasper

Logical Analysis and History of Philosophy

Philosophiegeschichte und logische Analyse

18

Analytic Philosophy Meets Classical German Philosophy

Guest Editors / Gastherausgeber
Birgit Sandkaulen · Albert Newen

mentis
Münster

Kant and Cognitive Science Revisited

Tobias Schlicht, Albert Newen (Bochum)

Abstract

To which extent is it justified to adopt Kant as a godfather of cognitive science? To prepare the stage for an answer of this question, we need to set aside Kant's general transcendental approach to the mind which is radically anti-empiricist and instead turn our attention to his specific topics and claims regarding the mind which are often not focus of Kant's epistemological investigations. If someone is willing to take this stance, it turns out that there are many bridges connecting Kant with contemporary cognitive science. We investigate possible bridges suggested in the literature between some of Kant's central claims about consciousness, mental content, and functions of mind, and some specific treatments of these topics in contemporary philosophy of mind and cognitive science. While doing so, we offer additional arguments for some proposed bridges, reconstruct others and completely destroy still other bridges by demonstrating that some suggested links between Kant and cognitive science remain only apparent.

Introduction

In *Kant and the Mind*, Andrew Brook (1994, 12) argues that Kant has been "adopted as an intellectual godfather by cognitive science". This is a surprising claim, given Kant's general anti-empiricist transcendental approach, in contrast to the empiricist approach of cognitive science. Methodologically, it seems that David Hume might be a more likely candidate for the title of "intellectual godfather of cognitive science" (Biro 2004; Newen 2003; Jacobson 2007). Yet, according to Brook, classical cognitive science has adopted several of Kant's central claims about the mind, most notably the claim that "most representations require concepts as well as percepts", and one central aspect of his method of transcendental argument, understood as the attempt to "reveal the conditions necessary for some phenomenon to occur" (Brook 1994, 12). Moreover, Brook interprets Kant's agnosticism about the underlying substrate of the mind in terms of contemporary functionalism, arguably the philosophical foundation of cognitive science (Boden 2007). According to functionalism, mental phenomena are exhaustively defined by their causal role or function, such that they can be described wholly independently of any commitments to the (physical) implementation of these functions. One could add to this list Kant's emphasis of the unity of consciousness, which has been rediscovered more recently by philosophers and cognitive neuroscientists alike (Bayne 2010). Obviously, if all these aspects of Brook's interpretation of Kant's philosophy of mind could be justified, his main assessment would be convincing.

In this paper, we do not intend to dwell on this very general point, but will focus instead on specific topic-related claims in Kant's theoretical philosophy, in order to evaluate the possible bridges between Kant's philosophy and contemporary cognitive science. Thus, we shall address only briefly the issue of functionalism and the metaphysical background of the mind-body problem. Brook is not alone in his functionalist interpretation of Kant's

putative philosophy of mind (Sellars 1974; Meerbote 1989). The main problem for this interpretation is that contemporary functionalism is not entirely ontologically neutral: it has been proposed as a reductionist theory, since the analysis of mental phenomena in terms of their causal roles is usually complemented by an additional claim about (possibly multiple) *physical* realizations of these mental functions (Kim 2005; Levine 2001; Chalmers 1996; Block 2015). In this sense, Kant would clearly oppose functionalism. It is questionable whether all the "functions" of the mind (CPR, A78f/B103f) that Kant invokes can be functionalized in the sense of being "realized" by a physical mechanism (Allison 1996). Thus, it is not surprising that a number of alternative interpretations of Kant's stance on the mind-body problem have been put forward, e. g. mere immaterialism (Ameriks 2000), dual-aspect theory (Sturma 1985), and epistemological dualism (Schlicht 2007). Whatever one makes of these diverse interpretations, with respect to this fundamental metaphysical question it is important to keep in mind Kant's transcendental idealism, which he himself calls "a dualism" (CPR, A370). It is certainly an understatement to say that Kant's transcendental idealism, i. e. the distinction between appearances and things in themselves, is not very popular among analytic philosophers of mind and cognitive scientists. The dominant background assumption guiding philosophical and empirical investigations of mental phenomena today is in general naturalistic, and, in particular, involves a commitment to some (either weak or strong) version of ontological physicalism. Current theories are formulated within a completely different paradigm from Kant's theory. Thus, if the question is whether Kant's original transcendental account of the mind has any bearing on contemporary scientific investigations of the mind, then the answer would likely be no: but it nevertheless seems to be one of the most interesting alternative paradigms of thinking about the mind, especially given the explanatory shortcomings of physicalims (cf. the contrastive discussion of these paradigms e. g. in Newen, Vogeley 2008). But to simply ignore Kant's position on the basis of his anti-empiricist stance would be much too quick, for it is not necessary to adopt Kant's general epistemological paradigm in order to appreciate that many of his specific claims are very much alive today. This is partly because he discovered important features of the mind that any satisfactory theory of mental phenomena must respect and take into account. In this paper, we would therefore like to bracket Kant's transcendental idealism, and instead discuss specific Kantian claims about the mind and evaluate whether close equivalents of these claims feature in contemporary debates in cognitive science. This leaves room for the view that Kant's work is an important source of some modern ideas, and may even inspire some modifications of modern theories. Given that we are setting aside some central aspects of his epistemological framework, we cannot claim to offer a thorough exegetical perspective, but we nevertheless aim to shed new light on some Kantian claims. This approach is also motivated by the fact that Kant never formulated a full-fledged philosophy of mind, such that what he says about these specific topics is scattered throughout his work, always in the service of another project, e. g. the investigation of the foundations and limits of knowledge. There has been huge progress since Brook's publication on various fronts. Needless to say, cognitive science has discovered many interesting phenomena and produced numerous theories about a vast range of mental phenomena. But, although Kant's views haven't changed (of course), philosophers have proposed many new and often surprising interpretations of his views based on his publications. Thus, we take the questions of how Brook's assessment fares today and, more specifically, how specific topics discussed by Kant are treated today to be relevant and interesting.

The issues that we intend to pick up in this paper are the following:

(1) Kant is often associated with 'conceptualism' about perception, i. e. the claim that perceiving involves the application of concepts (McDowell 1994). This is interesting for two reasons. First, this apparently intimate interrelation between percepts and concepts is central to the contemporary debate on 'cognitive penetration' (Stokes 2013). Secondly, it has recently been argued that Kant should be interpreted as defending what is nowadays called non-conceptual content with respect to perceptual experience (Hanna 2005).
(2) Kant famously claims that for a mental representation to be something for *me* it has to be possible for the 'I think' to accompany that representation. Proponents of the higher-order theory of consciousness often mention Kant as an inspiration for their theories (Gennaro 2004), but analysis of Kant's claim suggests that taking it to amount to a higher-order account is questionable. However, we intend to show that his fragment of a theory can be developed into an interesting alternative to present-day versions of representationalism (e. g. Tye 1995; Rosenthal 2005).
(3) A related claim is that object representations presuppose the contribution of intrinsic features of the mind, i. e. a faculty of spontaneity, which structures and regulates sensual input. It is interesting to note that neuroscientists have recently drawn a connection between their investigations of certain features of brain activity (e. g. chaotic bursts of self-generated neural activity on the one hand, and resting state activity on the other) and Kant's notion of spontaneity (Hanna & Thompson 2003; Northoff 2012). We will critically evaluate these claims.
(4) Kant's claim that we cannot understand organisms with mental capacities by mere mechanistic explanation, but must instead invoke teleological explanations has been explicitly picked up by Francisco Varela, in his elaboration of what has since been called "autopoietic enactivism" (Weber & Varela 2002; Hutto & Myin 2013), a striving new philosophical paradigm in cognitive science. This connection is interesting with respect to the project of naturalizing the mind, and will be discussed at the end of the paper.

1. The role of concepts in Kant's philosophy

1.1. Non-conceptual content

When he claims that Kant is the godfather of cognitive science, Brook appeals to Kant's thesis of an intimate connection between percepts and concepts, a thesis that has been taken up in cognitive science. Mole (forthcoming) suggests that Kant's claim that perceptual experiences of objects in the world involve a complex interaction between (the spontaneous application of) *concepts* and (passively received) *sensual input* can be translated into the recent debate about *cognitive penetrability*. Mole suggests that "the sort of interpenetration of perception and cognition" important in this latter debate

> underpins an important Kantian insight: that the world as we perceptually encounter it may be intelligible to us partly because the constraints of intelligibility play a role in determining the form of that very encounter. (Mole, forthcoming)

That is, according to a popular interpretation of Kant's epistemology, spontaneity as our capacity for intelligent thought is always already involved in the process of perceptual uptake. This is reflected in Brandom's statement that

> one of Kant's epoch-making insights … is his recognition of the *primacy of the propositional*…, the claim that the fundamental unit of awareness or cognition, the minimum graspable, is the *judgement*. (Brandom 2002, 510)

One question that may be raised is whether Kant defended a strong version of what is now called "cognitive penetrability". For, if Kant indeed claims that content is always propositional, then the content of our perceptual experiences cannot be non-conceptual, i. e. independent of (or impenetrable by) non-perceptual cognitive processes. As the discussion below will reveal, an assessment of this issue is more difficult than Brandom's unequivocal quote above suggests, and Mole's assessment also needs to be cashed out in a more differentiated way.

First of all, Kant of course famously held that "thoughts without content are empty, intuitions without concepts are blind" (CPR, A51/B75). In other words, "neither concepts without intuition corresponding to them in some way nor intuition without concepts can yield a cognition" (CPR, A50/B74). On the basis of this claim – that cognition always rests on both passive receptivity of sensual information *and* on the active application of concepts – most interpreters have concluded that Kant takes perceptual experience to be conceptual through and through. Thus, it follows that there is no such thing as non-conceptual perceptual content (McDowell 1994). But we should distinguish between what counts as *knowledge* and what counts as *conscious experience*. Even though the former may be conceptual, it does not follow that *all* conscious experience must be conceptual. In this vein, Hanna (2005) has argued that the situation is more complex, since Kant can be held responsible for 'conceptualism' and also for the contrary position, 'non-conceptualism'. According to the latter position, perceptual content can simultaneously be non-conceptual and cognitively significant. Different creatures can represent the world in many ways, either if they possess no conceptual capacities or if they are not able to apply the conceptual capacities they have in a given situation. This point is usually attributed to Gareth Evans (1982). In the context of perception, Evans and others have argued that our conscious perceptual states are so rich in phenomenal character that we cannot possibly bring to bear enough concepts to capture such states in all their richness. Furthermore, we can visually discriminate many more colors and hues simultaneously visible in one situation than we can recognize independently in different situations. This is called the fine-grainedness observation supporting the non-conceptuality claim of visual content (Raffman 1995). By accepting non-conceptual content, it is also possible to ascribe perceptual experiences to nonhuman animals and young infants who are either not yet or never to be in possession of the relevant conceptual abilities.[1] To what extent does Kant account for non-conceptual content?

Although Kant clearly states that any cognition always involves the combination of receptivity (sensibility) and spontaneity (see below), he also claims that something can appear to us in intuition without any function of the understanding ("Verstand") being involved:

---

[1] For further discussion of non-conceptual content, see Bermudez & Cohen (2008) and Gunther (2003).

> Objects can indeed appear to us without necessarily having to be related to functions of the understanding. (CPR A89/B122)

One way to reconcile these two claims is to distinguish understanding ("Verstand") from spontaneity (see also Kitcher 1990): "Verstand" is defined as the ability to conceptualize, and is thus clearly a conceptual ability. But spontaneity is a more general notion, as we will see below, and can be characterized as the general ability of unification ("Synthesis"). Even an appearance of an object without concepts being involved (what Kant calls the given manifold) needs some (minimal) unification of material by our senses. Thus, in line with Robert Hanna's interpretation, we can read Kant such that there is a possibility for non-conceptual content in intuition that, whilst not involving concepts, nevertheless involves a low-level unification ("synthesis"). In a footnote (CPR, B160/161), Kant seems to describe just this case of non-conceptual synthesis:

> Space, represented as *object* (as is really required in geometry), contains more than the mere form of intuition, namely the *comprehension* of the manifold given in accordance with the form of sensibility in an *intuitive* representation, so that the *form of intuition* merely gives the manifold, but the *formal intuition* gives unity of the representation. In the *Aesthetic* I ascribed this unity merely to sensibility, only in order to note that it precedes all concepts, though to be sure it presupposes a synthesis, which does not belong to the senses but through which all concepts of space and time first become possible. For since through it (as the understanding determines the sensibility) space or time are first *given* as intuitions, the unity of this *a priori* intuition belongs to space and time and not to the concept of the understanding.

Hanna argues that Kant allows for the non-conceptuality of an appearance, and attempts to provide a general positive characterization of non-conceptual content by relying on Kant's analysis of the structure of perceptual experience, which he locates in Kant's theory of the forms of intuition, namely, our representations of space and time, which are supposed to "*constitutively explain* non-conceptual content" (Hanna 2005, 278). More precisely: "Non-conceptual content is *nothing but* cognitive content that is essentially structured by our a priori representations of phenomenal space and time" (ibid.). In short, with respect to the quote above, it can be said that spatiotemporal structure or *form of intuition* organizes and pre-formats all perceptual content and locates perceived objects in space and time:

> The forms of intuition constitute non-conceptual content by introducing designated intrinsic phenomenal spatial or temporal structures into all human or non-human sensibility, whose specific cognitive-semantic function it is to determine the empirical representation of individual material objects in real empirical space and real empirical time, by uniquely locating those objects. (Hanna 2005, 282)

If Hanna's analysis is correct, then the predominant interpretation of Kant's theory of perception proposed by McDowell and others seems far too simple, and perhaps even wrong. Of course, it remains true that, for Kant, both intuition and concept are necessary for *knowledge*, but to have a conscious experience it seems sufficient to have an intuition that involves non-conceptual content structured acc. to space and time. According to this interpretation, it does not follow that all our perceptual experience is conceptual through and through, since we can have non-conceptual appearances. These can be characterized in modern terminology as conscious perceptual experiences. This general line of argument can be further supported by looking at Kant's characterization of aesthetic experience.

1.2. Aesthetic experience can be non-conceptual

In his discussion of the features of *aesthetic* experience and appreciation, in the *Critique of Judgment,* Kant analyzes our judgment "x is beautiful". Reflective judgments occur when we are intentionally directed at a specific object, either perceiving it or imagining it, and try to bring it under a specific concept, yet *lack* the concept needed in this situation. When we come to *know* something, the faculty of understanding provides a concept, which we can ascribe as a predicate to an object by judging, for instance, that "this is a car". But if we do not yet know the concept under which we could subsume the object in question, then our faculty of reflective judgment comes into play by looking for a general term that would help to classify what we imagine or perceive. In this process, the faculty of understanding (being responsible for concepts) and the faculty of imagination (being responsible for producing a unified intuition out of a given sensory input) may spontaneously work in full *harmony*. Imagination and understanding working together to constitute knowledge may then establish a state of "free play". The imagination is not constrained by a concept, but the unified representations it produces are not chaotic. We simply represent something imaginatively or perceptually without possessing a concept that would enable us to conceptualize it. Since this is, according to Kant, a state we appreciate, we are prompted to judge that the object or whatever is represented, which we are currently unable to classify with any specific concept, *is beautiful*. But we do not thereby recognize a property of the object, i. e. the judgment "x is beautiful" merely has the same form as a predicative judgment. What really happens, according to Kant, is that we judge our *representation* with respect to our current state of mind: we appreciate the free play of imagination and understanding for which we hold the represented object responsible (see Allison 2001 for details and discussion). The upshot of this brief characterization is that the conscious experience we enjoy in such a state has content, and presents an object of some sort, but since in this state of aesthetic appreciation *we lack the concept* to properly determine the object, the content of this representation must be characterized as *non-conceptual*. This is an additional way in which one can argue in favor of non-conceptual content within Kant's framework. Now, if Kant can allow for non-conceptual perceptual content in principle, then the dominant interpretation of Kant's view that every experience is conceptual associated with McDowell is wrong, and offers a far too simplistic picture. Brandom's assessment of Kant's view as one that asserts the overall primacy of the propositional with respect to awareness is then unjustified.

1.3 Clarifying the role of cognitive penetration in Kant's work

Let us now examine a further modern claim associated with Kant, i. e. the claim that perception is "cognitively penetrated." As mentioned above, Mole (forthcoming) interprets Kant as saying that perception is *always* so penetrated. This contradicts the result of the foregoing sections, namely that Kant allows for nonconceptual representational content of experience. Thus, we need a qualification of Mole's general claim.

We can take a first shot at defining cognitive penetration by relying on the opposite notion. According to Macpherson,

> perceptual experience is cognitively impenetrable if it is not possible for two subjects (or one subject at different times) to have two different experiences on account of a difference in their

> cognitive systems which makes this difference intelligible when certain facts about the case are held fixed, namely, the nature of the proximal stimulus on the sensory organ, the state of the sensory organ, and the location of attentional focus of the subject. (MacPherson 2012, 29)

To formulate this in positive terms, a case of cognitive penetration of a perceptual experience demands that, assuming a fixed sensory input, attentional focus, and state of the sensory organ, our *experience is modified* by the influence of a cognitive state (or process) like belief, desire etc., compared to a perceptual experience without any intervening cognitive states (or with different intervening cognitive states). It is important for cognitive penetration that the perceptual *experience* is modified, not the *judgment* based on this experience.

The cognitive penetration debate in the last two decades has focused on the claim that perceptual experiences, more precisely *early perceptual processes*, are *always* cognitively impenetrable (Fodor 1983; Pylyshyn 1999). To illustrate and argue for this claim, Fodor and Pylyshyn rely on the Müller-Lyer illusion, according to which we perceive two straight lines to be different in length *even* if we *know* (perhaps because we have measured them) that the lines are of equal length. This view implies that although early perceptual processes influence our beliefs and other cognitive states through bottom-up processing, there is no top-down influence from beliefs and other cognitive states to these early visual processes. The latter are characterized as forming a module which is informationally encapsulated from, i. e. cannot be influenced by, cognitive states. This claim has been challenged recently by a growing group of philosophers and cognitive scientists (e. g. MacPherson 2012; Siegel 2012; Stokes 2013; Lupyan 2012). Given this background, the thesis about the cognitive penetrability of perceptual experience is a modest claim, i. e. it says only that perceptual experience is *sometimes* cognitively penetrated. It is *not* the claim that perceptual experiences is *always* cognitively penetrated.

Concerning this modest understanding of cognitive penetration, it is clear but not astonishing that Kant can be categorized as defending cognitive penetration. In the case of perception-based *knowledge*, our perceptual experience of a city map *as* a city map is shaped by concepts. Not only the *judgment* that this is a city map depends on concepts; the activation of the concept of a city map also modifies the perceptual *experience*. We can ascribe to Kant the position that we would have a different *experience* when looking at the same entity (the map) if we did not conceptualize it as a city map but e. g. as a modern painting of brain structures (or if we had no concepts available). The disputed question is whether, on Kant's view, all perceptual experiences are cognitively penetrated. Is he defending what we may call radical cognitive penetration?

In order to answer this question, we can refer directly to our earlier results. Cognitive penetration demands the influence of a cognitive state on early perceptual processes (given fixed background conditions, see above). If the relevant penetrating cognitive states are narrowly defined as *conceptual* states, then Kant does *not* hold that *all* perceptual experiences are cognitively penetrated, since we have seen that he allows for non-conceptual experiences. But if we broaden the notion of penetrating cognitive states (see Vetter & Newen 2015) to include any process of combination ("synthesis"), then Kant defends radical cognitive penetration, because on his view *all* perceptual experiences involve a process of combination; it is just that the act of combination need not always be executed by the understanding.

But defenders of cognitive impenetrability such as Pylyshyn allow that intra-perceptual processes in a given module may involve processes of combination. Thus, to characterize Kant as defending radical cognitive penetration on the basis of the role of "synthesis" would be systematically inadequate, because it would undermine the original definition of cognitive penetration.

To summarize the relation between Kant and cognitive science concerning the role of concepts, we have argued 1. that Kant allows for non-conceptual experiences that involve a process of non-conceptual synthesis, 2. that aesthetic experiences can be non-conceptual according to Kant and give additional support for the first thesis, and 3. that we can readily attribute to Kant the standard claim of cognitive penetration, but that it would be inadequate to attribute a radical claim about the cognitive penetration of perceptual experience to him.

The next contemporary debate that shall be taken up and compared to Kant's account concerns the question of what marks the difference between those representations that are consciously experienced and those that are not. What is the condition that the former meet that the latter do not?

2. The mineness of conscious experiences

Before we turn to an analysis of Kant's characterization of the mineness of conscious experience, it will be useful to clarify the notion of subjective experience in contemporary debates and the predominant approach to it, namely representationalism. A central discussion in contemporary philosophy of mind concerns the difference between phenomenally conscious and unconscious mental representations. In the last two decades, numerous proposals have been formulated by both philosophers and empirical scientists, most notably cognitive neuroscientists. The relevant sense of consciousness here is consciousness in the sense of what Nagel called subjective character: "The fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism" (Nagel 1974, 436). A creature is phenomenally conscious if it has an experiential point of view, i. e. if there is something that it is like, subjectively, to *be* that creature. A mental representation may be regarded as phenomenally conscious if and only if there is something it is like for an organism to *be* in that state (or to *have* the experiential state in question), i. e. if the state is something *for* the organism. The mental representation is then supposed to exhibit phenomenal character. Unconscious representations are those for which there is *not* something that it is like *for* the organism. Typical examples of phenomenally conscious experiences are sensations such as pain, and perceptual experiences (for instance visual or auditory experiences). Typical examples of representations that are unconscious in this sense are blindsight or priming (Marcel 1983). But there is also a debate about what is sometimes called "cognitive phenomenology", i. e. the question of whether there is also something it is like to have propositional attitudes like beliefs (Bayne & Montague 2011).

2.1. Representationalisms

What is the condition that phenomenally conscious representations meet but unconscious ones do not meet? Conceptual and empirical answers have been given to this question. The dominant group of philosophical theories are versions of representationalism. Their

core assumption is that phenomenal consciousness can be analysed in terms of the notion of representation or intentionality, i. e. the mind's capacity to be directed at something or be about something. The different versions of this position then differ with respect to how they take the notion of representation to play a central role in an explanation of the subjective character of phenomenally conscious representations. For example, according to Tye's (1995) first-order representationalism, a mental representation is phenomenally conscious if it has abstract, non-conceptual intentional content that is "poised", i. e. available for further use by the conceptual system yielding beliefs and judgments. According to higher-order versions of representationalism, this condition on first-order representations is not sufficient. Instead, a mental representation must itself be represented by yet another representational state, for instance a thought, to the effect that 'I am now having this representation' (Rosenthal 2005). To avoid the problems that arise from introducing this second representation, a third version has been proposed, holding that a mental representation is phenomenally conscious if it not only represents its object, for example the blue sky, but also *itself* (Kriegel 2009). All these proposals have been discussed at length elsewhere (Gennaro 2004).

Now, it is interesting to note that proponents of higher-order theories of phenomenal consciousness, like Lycan, Rosenthal, and Gennaro, explicitly rely on Kant's notion of inner sense to explicate their idea that for one of my mental representations to be conscious is for *me* to be conscious *of* that representation (e. g. Gennaro 2004, 3). So, the next step is to see whether Kant's (fragments of a) theory actually fall into the same category as these views, or whether they can guide us in developing an alternative.

2.2. Kant on mineness

In Kant's view, conscious representations only form a small subset of representations in general. Among all representations, Kant distinguishes – in the form of a 'progression', as he puts it – unconscious from conscious representations, and among the latter between merely subjective and objective representations:

> The genus is representation in general (*repraesentatio*). Under it stands the representation with consciousness (*perceptio*). A *perception* that refers to the subject as a modification of its state is a sensation (*sensatio*); an objective perception is a cognition (*cognitio*). The latter is either an intuition or a concept (*intuitus vel conceptus*). (CPR, B376f)

What, in Kant's view, marks the difference between conscious and unconscious representations? His central and famous claim is that

> the *I think* must *be able* to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least be nothing for me. (ibid., B131)

If a mental representation cannot possibly be accompanied by the 'I think', then there is nothing it is like *for me* to have it, according to Kant. This may happen for two different reasons. First of all, a representation is *impossible* if it is self-contradictory, i. e. if it cannot be thought at all. In this case, it remains "nothing for me" for *logical* reasons. Thus, it is impossible for me to form a representation of a circular triangle. Secondly, a representation can also be *nothing for me* for reasons that have nothing to do with its content, namely for *cognitive* reasons. *Obscure* representations remain unconscious because I cannot even know

that I have them. So-called subliminal perceptions, as we know them from patients with blindsight or from priming experiments, are a good candidate for obscure representations. Although they do not reach the threshold of conscious experience, they do have behavioural effects within the subject's total mental economy. Kant allows for an "immense" area of such obscure unconscious mental representations in his *Anthropology* (Kant 2007, cf. Kant AA VII, § 5, 135f).

Obscure representations are contrasted with "clear" representations. These can be accompanied by the 'I think' and can consequently be something *for me*. A representation is "clear" insofar I am conscious of its content and can distinguish it from other representations (cf. CPR, B415n.). The 'I think' functions as an act or a process which presupposes sensory (or other) representations given in inner sense, which in turn depend, in order to become conscious (something *for* the subject), on the execution of this act. So this is a kind of mutual dependence, since the latter would "not take place" if no "material" in inner sense was presented (CPR, B422n.).

What's important here is that the 'I think' – the unity of apperception – understood as an "act" is contrasted with "inner sense", the general capacity to have representations at all and become aware of them (ibid., B153). All representations, stemming from the outer senses, the imagination, dreams etc., are contained in inner sense, and ordered in accordance with temporal relations. According to Kant, inner sense presents a given "manifold" of obscure representations "without spontaneity" (ibid., B68). This manifold is constituted by such obscure representations that are not yet clearly distinguished from one another. They *can* be apprehended and thus be clearly distinguished from each other, i. e. they can be made *conscious*. But they only possess the *possibility* of becoming conscious. This requires a spontaneous activity on the part of the subject called "self-affection" (ibid., B67ff, B153ff). That is, the subject of experience can *affect* itself, i. e. its inner sense. What is interesting here is that, according to Kant, "attention" is an example of such an act of self-affection (ibid., 156f n.): attention clearly sorts out the sweeping representations from each other and combines some of them in the unity of a thought. In this case, our sensibility is affected by our spontaneity or understanding. Such an act of self-affection can thereby illuminate the contents of (and the differences among) various representations, and order them in specific (e. g. temporal) relations. In short, the inner sense, which is also called the "passive subject" (ibid., B153), is affected by the "faculty for becoming conscious of oneself" by apprehending that "which lies in the mind" (ibid., B68). This spontaneous self-consciousness, represented by the 'I think', cannot *produce* representations itself, since it depends on material from the various forms of sensibility given in inner sense. But it is through attentional focus that we can become conscious of a representation by clearly apprehending its content. Obviously, these aspects of Kant's philosophy need to be elaborated in much more detail, but for reasons of space, this sketch must suffice (see Nakano 2011, Pippin 1987 for details). The result is a threefold distinction between representations that *are* accompanied by the 'I think', those which merely *can* be so accompanied, and those which *cannot* be so accompanied.

2.3. Kant and the Global Workspace: an analogy

It is tempting to compare Kant's classification with the threefold distinction between unconscious, pre-conscious and conscious representations proposed by Dehaene et al. (2006). The central idea of their neuro-functionalist model of consciousness, the so-called global workspace, is that the function of consciousness is to make integrated information globally available for the control of action and for verbal report (Baars 1988). Dehaene's team has developed Baars' purely functional model further with respect to its neural implementation in the brain, suggesting that the workspace comprises structures in the prefrontal cortex and in general frontal areas of the brain, which are often associated with *cognitive* functions. The first condition for a representations' being conscious is that the object stimulus must mobilize excitatory neurons with long-range connections across the cortex, and thus be able to support global activity in the brain. This neural firing supporting any conscious object representation must reach a minimal strength. At any moment, many object representations compete for access to consciousness but only some can be consciously experienced at any given moment. On the neural level, the respective neural correlates of these representations likewise compete for access to the global workspace. The relevant filter is attention, which, according to this theory, is a *necessary condition* for conscious experience, such that we are only aware of what we attend to. This of course seriously restricts the capacity of consciousness to the capacity of attention. This point is the subject of ongoing debate, since critics argue that the capacity of consciousness "overflows" the capacity of attention and thus for cognitive access and report (Block 2011).

Within Dehaene's model, this leads to a threefold distinction among neural representations: (a) Representations in the focus of attention, and whose underlying neural activation exhibits the relevant strength, are integrated into the global workspace and count as *conscious*. (b) By contrast, representations outside the focus of attention whose underlying neural activation has nevertheless reached the minimum strength are called *pre-conscious*. These are neither something *for* the subject nor cognitively accessible for use in reasoning or speech. But a shift of attention may integrate such representations into the global workspace and thereby make them conscious. (c) Finally, representations whose underlying neural activations lack the relevant strength eventually fade out and cannot be amplified by attention: these are and forever remain *unconscious*.

On Kant's taxonomy, conscious representations that are something *for me* are the ones we are conscious of, in virtue of being (possibly) accompanied by the 'I think': (a) A representation *actually* accompanied by the 'I think' corresponds to a representation that is integrated into the global workspace (via attention). This secures cognitive access to this representation, which is consequently something for *me*. (b) What Dehaene calls pre-conscious can be taken to be the set of representations that *can be* but are *not actually* accompanied by the 'I think', i. e. representations which *can* be integrated into the global workspace. (c) What counts as an unconscious representation in Kant's theory – not being possibly accompanied by the 'I think' – also counts as unconscious in Dehaene's framework. These are the representations that can never be integrated into the global workspace. The interesting question is whether there is a match regarding the second group of representations presented under the (b) category.

Pre-conscious representations are not currently cognitively accessed, yet are accessible in principle. However, in Kant's framework, this is not so clear cut. According to Heidemann's

(2012) helpful analysis, these representations are "clear" yet merely "indistinct" (but not "obscure" in Leibniz's sense), in contrast with the clear and "distinct" representations that *are* accompanied by the 'I think'. Heidemann argues that Kant's distinction between distinct and indistinct representations is orthogonal to the distinction between sensibility and understanding. More specifically, on his reading of Kant, "clear and distinct representations are conscious *simpliciter* whereas clear but indistinct representations are conscious by degrees". More specifically, "indistinct representations [...] are both unconscious by degrees and conscious by degrees" (ibid., 50). These latter representations are cognitively accessible, since it is *possible* for the 'I think' to accompany them, and this possibility alone ensures cognitive access. Thus, since Kant allows for representations "that are not per se thought by the 'I think'" (ibid.), i. e. representations that we are not *explicitly* aware of, he seems to make a distinction between conscious experiences that are phenomenal and cognitively accessed, and experiences (or rather, representations) that are only possibly accessible. On Kant's view, then, for representations to be *my* representations, they need not be representations I am conscious *of*. This seems to contrast with Rosenthal's higher-order theory, and to be more in line with Block's (2011) conception of phenomenal representations that are something *for me*, yet only accessible under certain circumstances (cf. Schlicht 2012).

Heidemann draws a connection between this aspect of Kant's theory and the possibility of non-conceptual representations, such that indistinct intuitional representations are non-conceptual. According to Heidemann's reading of Kant, there are aspects of human cognition that are neither rational nor conceptual. As an example, Heidemann mentions

> phenomenal consciousness. A person perceiving a rainbow might not be distinctly aware of a certain color; but by means of the "I think" she can bring to her mind or become self-conscious that she herself sees this particular color shade although she might not be able to conceptualize what it is like to see it. (ibid., 57)

Thus, it seems that although we can match Dehaene's thoroughly empirical threefold taxonomy of conscious, pre-conscious, and unconscious representations to Kant's threefold taxonomy, there is room for debate whether the representations that *are not currently*, but *can* be accompanied by the 'I think' count as being something *for* the subject or not. This is due to Kant's relatively weak condition of mere *possible* accompaniment by the 'I think'. Yet, even in Kant's model, this merely logical representation has an empirical correlate, my *unified* phenomenal experience. The unity of consciousness is closely associated with the 'I think', of course, and the following section investigates this connection and the possibility of this leading the way towards an alternative to the various versions of representationalism, especially the higher-order theory of consciousness.

2.4. Consciousness and Integration

Kant is famous for having stressed *unity* as a central feature of conscious experience. Indeed, according to Brook (2007), Kant took this to be the most important feature of consciousness. Another way of saying that it has to be possible for the 'I think' to accompany all my representations is to say that it must be possible for a manifold of diverse representations to be contained in *one* and the *same* consciousness, since otherwise "they would not altogether be *my* representations" (CPR B132), i. e. there would not be something that

it is like *for me* to have them. But in Kant's parlance, this is equivalent to saying that a mental representation is conscious if it is *integrated* into *the one global representation* we experience, along with other representations that make up its parts. There is only one unified global representation that makes up my *total state* of consciousness at any moment (see Brook 1994, 80ff). In Kant's theory, the empirical phenomenal unity of consciousness, i. e. the one we *experience*, is due to the transcendental unity of apperception, as signified by the 'I think' from which all empirical unity emanates. Nevertheless, the synthesizing act 'I think', understood as integration, does not refer to an ontological entity that can be said to exist (or take place) *independently* of given sensory input: "without any empirical representation, which provides the material for thinking, the act I think would not take place …" (B422n.). Consequently, in an analysis of the 'I think', it is wrongheaded to look for an ontological entity; this becomes clearer if we simply translate the accompaniment by the 'I think' as the integration into the unity of consciousness. Relatedly, the self-consciousness that is made possible via a conscious representation is not a consciousness of self as *object*, but as *subject*: When I am consciously seeing the blue sky, then the representation of the sky is conscious in virtue of being integrated into the global representation I am enjoying right now, and thus in virtue of being accompanied by the 'I think'. If a representation is not so integrated, then it is *nothing for me* – it remains unconscious. If it wasn't for the spontaneous act of synthesis (or integration), a disorganized "swarm of appearances" (B134) could "fill up our soul" (A111) and produce as "multi-coloured, diverse a self" as there are representations. Thus, our experience is only *one*, since all *my conscious* representations belong to (are integrated into) the same consciousness in virtue of an act of integration or binding, performed by the subject's power of understanding.[2] In this way, the experienced unity of consciousness and the 'experiencable' unity of self are intimately related.

In current debates, many theories are versions of representationalism (e. g. Tye 1995; Rosenthal 2004), according to which a certain form of *content* or representational relation is held responsible for the difference between conscious and unconscious representations. Opponents of such views (e. g. Vosgerau et al. 2008) may hold that although conscious experiences in general *represent* something or other, it is not because of some *representational* relation that they are phenomenally conscious (or exhibit subjective character). Proponents of such alternative theories deny that any specific *content* is constitutive of a conscious experience. Rather, it is some *process* or other that makes the important difference between a conscious and an unconscious representation (cf. Marchi & Newen 2015). Various alternative processes could be responsible for the difference between phenomenally conscious and unconscious representations. For example, a currently popular proposal is that it is relevant whether the mental representation is *integrated* in the right way into a global representation, i. e. if it is an *element* or *modification* of a much more encompassing conscious state. For example, Searle (1992) and Bayne (2010) emphasize 'phenomenal unity' as an important feature of consciousness. A representation of something red does

[2] Recently, Bayne (2010) also defended a version of Kant's claim that we *always* enjoy one single unified experience. Bayne distinguishes between various forms of unity pertaining to consciousness, for example access unity and representational unity, but argues that these can all break down under certain conditions, e. g. in split-brain patients, except the phenomenal unity of conscious experience. Our phenomenal experience is always unified, consisting of visual, auditory and other perceptions, sensations, thoughts, and memories.

not become conscious individually, but always in concert with other representations in a unified way. In the same vein, Van Gulick (2004) suggests that

> transforming a non-conscious state into a conscious one is a process of *recruiting* it into a *globally integrated complex* whose organization and intentional content embodies a heightened degree of *reflexive self-awareness*. (Van Gulick 2006, 24)

In other words,

> the transformation from unconscious to conscious state is not a matter of merely directing a separate and distinct meta-state onto the lower-order state but of 'recruiting' it into the globally integrated state that is the momentary realization of the agent's shifting transient conscious awareness. (Van Gulick 2004, 76f.)

On such a view, a mental representation is phenomenally conscious if it *modifies* (at any given time) the complex (or cluster) forming the *single global representation* or momentary "total state of consciousness" of a subject interacting with her environment (Bayne & Chalmers, 2003). If a representation is not integrated in this way (like, for instance, the visual representation of a blindsight-patient), then it remains unconscious (Weiskrantz 1986). The central idea is that the conscious experiencing self enjoys one unified global mental representation made up of many elements, some of which are stable and invariant, while others are constantly changing in virtue of the interaction of the whole organism with objects and events in the environment. A virtue of this approach is that it can accommodate the observation that any single mental representation, such as visually representing the blue sky, is always only a *modification* of a much bigger overall conscious state of an experiencing organism; the visual experience of the blue sky does not become conscious in complete isolation. Another virtue of this theory is that integration as a process can be characterized in purely functional terms, and can in principle be realized by neural mechanisms in the brain. Consequently, several empirical accounts support this general idea. For example, both Engel and Singer (1999) and Metzinger (1995) propose that neural oscillations in the frequency range of 40Hz might be responsible for the *binding* of various elements to a unified representation on multiple levels of complexity, from the unification of elements to an object to the unification of representations into a global representation. This is just a selection of theories that make use of the process of integration, and for reasons of space, we abstain from discussing more: the general point will suffice for present purposes.[3]

Having outlined these approaches alongside Kant's, one can easily see that Kant's account of what it takes for a mental representation to be *something for me* is very different from the higher-order theories on offer. As mentioned above, proponents of higher-order accounts such as Gennaro and Rosenthal often refer to Kant when they introduce their theories. Yet Kant does not envisage an account along the lines of the higher-order theory, because he does not think that a representational relation between a target representation (of the blue sky say) and a separate representation is responsible for the former being

---

[3] It is important to note that the process of integration is especially popular in empirical approaches to consciousness. For example, Edelman and Tononi (2000) postulate a globally integrated state as the substrate of consciousness and subjectivity. Damasio (1999) attempts to integrate what he calls 'the sense of self' into the unified phenomenally conscious experience, by alluding to structures in the brain that monitor and regulate the overall state of the organism *for which* there is something that it is like to have a given experience.

conscious. The 'I think' does not *represent* the representation of the blue sky when it *accompanies* it. Understood in the sense of *integration*, it means that the representation of the blue sky is conscious in virtue of its being integrated into the global representation of the experiencing subject. So Kant's approach is much closer in spirit to *integration* theories, as put forward by Van Gulick and others. This is important, because Kant's theory is therefore not prone to all the objections against higher-order accounts. It remains to be seen how to fill in the details of an account that retains Kant's insights but moves beyond his transcendental background theory.

3. Spontaneity of Mind and Brain

3.1. Neural correlates of consciousness

One of the most extensive research programs in contemporary cognitive neuroscience is the search for the neural correlate of consciousness (Metzinger 2000). The neural correlate of consciousness (ncc for short) is usually "defined as the *minimal neuronal mechanisms jointly sufficient for any one specific conscious percept*" (Tononi & Koch 2008, 239). The idea is to identify the neural mechanisms (areas, processes or kinds of neurons) that can help distinguish a conscious sensation of pain from a conscious perception of seeing the blue sky, say. Although there has been impressive progress in this area over the last twenty years, the idea faces ongoing problems and shortcomings.

Given the definition above, many researchers have supposed that the right place to search for the neural correlates of consciousness is the cortex (e. g. LeDoux 1999; Craig 2009; Lamme 2006). Indeed, the various suggestions in the literature point to a central role of the larger thalamo-cortical system. Part of the idea is that consciousness is made up of conscious units – units for vision, for auditory experiences, for tactile experiences and so on – and that "all the different aspects of consciousness [. . .] employ a basic common mechanism or perhaps, a few such mechanisms" (Crick and Koch 1990, 277). Searle (2005, 154ff.) has criticized this approach, since it neglects the fact that the participants (human beings, monkeys or whatever) in the relevant experiments are always already *conscious*, albeit in a much more basic sense than the phenomenal sense associated with a specific conscious experience (of seeing the blue sky, say). That is, when the alleged neural correlate of seeing the blue sky is activated, the state of the organism does not change from unconscious to conscious. Rather, the overall state of the organism (that was already conscious before) is merely *modified* in a certain way.

A distinction between the *core* ncc and the *total* ncc of a given sensation can help to clarify this confusion. No researcher believes that activity in the fusiform face area, say, which has been supposed to be the ncc for face perception and recognition, is sufficient *all by itself* for a perceptual experience as of a face. Rather, this activity must be embedded in and suitably connected to the rest of the total ncc of this perceptual experience. That's where the notion of *integration* as elaborated above comes in.

The crucial question, of course, is how to identify the structures constituting the total ncc and differentiate the various functional contributions of its elements to the total unified state of consciousness at a given time (Bayne & Chalmers 2003). One important piece of evidence that speaks against the corticocentrism associated with the search for the *core* ncc

stems from children with hydranencephaly, whose cortex is absorbed and replaced with cerebrospinal fluid. In these children, the remaining pieces of cortex are also disconnected from the thalamus, which is supposed to play a central role with respect to conscious experiences. While such patients do not enjoy any perceptual experiences whatsoever – be they visual, auditory or in other modalities – they are not completely unconscious. On the contrary, they show signs of being awake and alert, and also exhibit strong emotional responses to familiar stimuli (Shewmon, Holmse & Byrne 1999). Merker (2007) argues that such patients are deprived of certain pieces of information, but not of consciousness; on the basis of such patients, he concludes that consciousness *as such* is not generated by structures in the cortex but instead is generated *subcortically*. Solms (2014), agreeing with Merker, argues that this approach goes back at least to Moruzzi and Magoun (1949) who trace the origin of consciousness back to activation of the upper brainstem (see also Damasio 1999). Lesions in the upper brainstem lead to coma and other conditions of unconsciousness. In Damasio's terms, patients with a lesioned brainstem lose wakefulness, and consequently also mind and self (Damasio 2011, 234). A corresponding lesion in the cortex with preserved brainstem may lead to locked-in syndrome, in which the patients are not unconscious, but conscious yet paralyzed (ibid.). In general, this research supports the notion that the necessary conditions for consciousness *as such* in the brain are to be found in evolutionarily older regions, which are widespread in the animal kingdom.

To specify further the notion of a total ncc, some researchers have recently drawn an important distinction between stimulus-driven (or extrinsic) neural activity and self-generated (or intrinsic) brain activity (Hanna & Thompson 2003; Northoff 2012; 2013; Fazelpour & Thompson 2015). The former can be defined as the neural activity that occurs *in response* to the organism's interaction with the world or specific stimuli, while the latter "occurs prior to the onset of extrinsic stimuli that form the basis for the subsequent contents of consciousness" (Northoff 2013, 726). While the intrinsic neural activity cannot therefore be considered a sufficient condition of consciousness, it seems that it can *predict* or at least modulate the neural, phenomenal, and behavioural effects of subsequent stimuli (Northoff et al. 2010; Sadaghiani et al. 2010). The authors mentioned above have associated this intrinsic activity of the brain (to be determined more specifically below) with what they independently call the "Kantian brain." What do they mean by this, and to what extent is the association justified?

3.2. Functionalism, Enactivism and Dynamical Systems

As mentioned above, the dominant paradigm theory guiding research in philosophy of mind and cognitive science has been a version of functionalism. A guiding assumption of functionalism is that all mental information-processing functions can be realized by a physical mechanism (presumably in the brain, but not necessarily limited to it). Thus, a functionalist understanding of mental *terms* is typically complemented by a materialistic view of the mechanisms executing these mental functions. Brook has argued that, given his functionalist interpretation of Kant's theory, "materialism fits remarkably easy into his overall theory" and that Kant's "observations and inferences concerning synthesis, unity, and self-awareness can quite easily be made to fit into materialist theories of mind" (Brook 1994, 15). This may seem surprising, given that the Kantian spontaneity of mind, properly understood and explained below, seems incompatible with materialism (or physicalism).

Allison, for example, is less optimistic. He argues that Brook's functionalist-materialist interpretation of Kant's theory of the mind cannot be right, since, in Kant's view,

> cognition must be conceived as more than an elaborate information processing procedure, one which begins with raw sensible input and ends with the relatively reliable products of the understanding (cognitions). [...] What is missing in such a picture of cognition (at least from the Kantian perspective) is precisely its self-conscious, apperceptive character. (Allison 1996, 63)

Is there a way of incorporating the idea of spontaneity, associated with apperception, within a broadly naturalist framework? What could an analogue notion from within contemporary cognitive science be? While in the classical functionalist framework there does not seem to be room for the kind of spontaneity that Kant attributed to the mind, Hanna and Thompson (2003) and Fazelpour and Thompson (2015) argue that a different paradigm of cognitive science, namely enactivist-embodied cognitive science, can accommodate Kant's notion of spontaneity much better.

Cognitive science has developed and undergone a paradigm shift from classical computer-type functionalism to a thoroughly embodied and enactive cognitive science (e. g. Varela et al. 1991, Thompson 2007, Chemero 2009). In this paradigm, the varieties of cognition are *not* conceived of as instances of mere logical problem solving, where the brain is conceived of as a computer with the mind as its software. Rather, it must be interpreted by taking into account the complex dynamical relationships of the brain to the body and to the organism's environment (Clark 1997). Enactivism comes in different varieties, but the gist of this web of ideas is that brain activity alone is not sufficient for cognition and consciousness, since these phenomena are *active* achievements of whole organisms, suitably situated in and coupled to (aspects) of their environment (Hutto & Myin 2013).

Thus, the brain is conceived of as a "complex, self-organised system with nonlinear dynamics" (Singer 2013, 616), which is capable of spontaneously generating patterns of activity that form the basis of mental phenomena. This spontaneous activity is sometimes associated with the kind of spontaneity Kant had in mind. Hanna and Thompson understand Kant's notion of "spontaneity of consciousness" in terms of "its inner plasticity and purposiveness" (2003, 133f.) which, in their view, has been completely neglected in debates on consciousness, despite its importance:

> We believe that subjective experience is partially constituted by its being at once underdetermined or uncontrolled by external influences (inner plasticity), and also self-determining or self-controlling (inner purposiveness). It is this dual subjective sense of inner plasticity and inner purposiveness that we mean to indicate with the term 'spontaneity' as applied to conscious experience. (Hanna & Thompson 2003, 137)

In short, they argue that the notion of phenomenal consciousness that goes back to Nagel's (1974) description of there being "something it is like to *be*" a certain organism should be *complemented* by there being "something it is like to *do*" (from the perspective of an organism). This is in line with the central claim among enactivists that all cognitive phenomena are *active* achievements. Thus, cashing out the spontaneity of consciousness, they provide a list of four aspects:

> the precise qualitative character of conscious states (1) is not determined by anything external to the conscious subject; (2) is self-generated; (3) is not self-generated by a prior conscious intention; and yet (4) can under some conditions be controlled by a conscious intention. (Hanna and Thompson 2003, 147)

The enactive conception of an organism as a self-organizing and self-producing system *includes* the mind as such a self-organizing and self-producing (sub-)system. Thompson calls this the claim that 'mind is life-like' and 'life is mind-like', i. e. that there is continuity yet different complexity among the organizational features of life and the organizational features of mind (Thompson 2007, 128). The question is whether Hanna and Thompson justifiably associate (or reduce) Kant's notion of spontaneity to the 'intrinsic' neural mechanisms investigated by neuroscience.

They argue that multistable figures, such as the Necker-cube or the famous duck-rabbit, which admit of different interpretations or ways of seeing them, corroborate the notion of the spontaneity of consciousness. The visual pattern importantly *allows* for different alternatives, without determining which one is seen at any moment. Given that the stimulus does not change, despite the changes in subjective experience, the latter must be treated as an achievement or a performance of the perceiver. In that sense, the stimuli do not drive the neural processing, but merely act "as boundary conditions on the autonomous (self-organizing) processes of order formation in perception" (Hanna and Thompson 2003, 146). These structuring processes occur unconsciously. Yet these examples merely indicate the phenomenon of what we called "cognitive penetrability" above (in section 1).

Hanna and Thompson also allude to Kelso's work on synergetics, according to which neural activity exhibits a "chaotic dynamics called 'intermittency' [...] which corresponds to a nearly periodic motion that is interrupted by occasional, irregular, and unpredictable bursts" (Hanna & Thompson 2003, 152). In a state like this, the brain's overall state oscillates between regular and irregular chaotic behaviour. On this view, the switch from one way of seeing a multistable figure to another way of seeing it is an expression of such chaotic bursts. They generalize this contribution of the 'chaotic brain' to all cases of perception, since "every percept is potentially unstable" (ibid., 153) in this sense. It is in this way that Hanna and Thompson intend 'spontaneity' to be understood. Before we compare this to Kant's notion of spontaneity, let us introduce a different notion of the "Kantian brain".

3.3. The brain's resting state

In a number of writings, Northoff (2010; 2012; 2013) has associated Kant's notion of the spontaneity of mind with the brain's intrinsic activity. But, in contrast to Hanna and Thompson, he refers to the activity of the brain in the so-called "resting state" which has also been associated with the brain's "default mode network" (Buckner 2012). This network comprises a number of related brain areas, known as the 'cortical midline structures'. Although it seems clear that the resting state activations interact with the stimulus-activations in the brain, the exact features of the resting state that are important here are largely unknown. Northoff (2012, 357) claims that

> this is where Kant's view of the mind may become useful. What Kant described as the mind's intrinsic features, providing order and regularity to the extrinsic stimuli from the world, could be attributed to the brain's resting state and its intrinsic features. More specifically, the brain's resting state activity may structure and organise stimulus-induced activity in such a way that the latter can be associated with consciousness, self, and spatiotemporal continuity.

Activations in the cortical midline structures (belonging to the default network) are highly correlated with the presentation of self-related stimuli (in contrast to non-self specific

ones) (see also Vogeley et al. 2001). Northoff combines this finding with the observation of high resting-state activity in these same regions, and tentatively concludes that self-specificity could be represented in resting state activity. The idea is that if self-specificity is an intrinsic feature of the resting state, and if it organizes the neural processing of extrinsic stimuli, then this could be taken as an assignment of self-specificity, with the effect that the representations of the extrinsic stimuli "are ultimately experienced as part of one's self" (Northoff 2012, 358).

Yet, as Northoff also notes, activations in the cortical midline structures have also been interpreted as a neural correlate of 'mindreading' or social cognition, i. e. our capacity for understanding other people's mental states, e. g. beliefs, desires, emotions etc. (Frith & Frith 2000). This relativizes the finding, in the sense that ascriptions of mental states to oneself (self-related) and others (other-related) are importantly different in function and significance with respect to ensuing actions etc., such that activations in this network may be taken as *neutral* with respect to this factor. This seems to be an open empirical question.

3.4. Kant on Spontaneity

Let's look at Kant's notion of spontaneity more closely, and see whether it can be associated with the various notions of a Kantian brain just outlined. The result of Kant's "revolution" in epistemology is the assumption of a synthesizing-function of the mind (attributed to the faculty of understanding), which is conceived of as the mind's original contribution to cognition. As is well known, in contrast to the *passivity* of receptivity, the function of integration or synthesis is an essentially *active*, i. e. spontaneous achievement of the mind. Yet although it plays a central role in Kant's theory of the mind, this capacity to organize and synthesize sensory stimuli in novel ways relative to those stimuli has not proved popular among cognitive scientists and analytic philosophers of mind. This may be partly due to the fact that recent philosophy of mind has focused on sensations and their qualitative features, rather than on cognition in the sense of conceptual understanding. More importantly, the paradigm view of cognition in the latter sense has been the classical computational-representational view of the mind (Fodor 1975), in which there is no room for the Kantian function of synthesis. Why is this so?

According to Kant, the object of cognition is not what is received via the senses, but the product of an intellectual "combination". It is a central assumption of his epistemological project that

> the *combination* (conjunctio) of a manifold in general can never come to us through the senses, and therefore cannot already be contained in the pure form of sensible intuition; for it is an act of the spontaneity of the power of representation, and, since one must call the latter understanding, in distinction from sensibility, all combination, whether we are conscious of it or not, whether it is a combination of the manifold of intuition or of several concepts, and in the first case either of sensible or non-sensible intuition, is an action of the understanding, which we would designate with the general title *synthesis* in order at the same time to draw attention to the fact that we can represent nothing as combined in the object without having previously combined it ourselves, and that among all representations *combination* is the only one that is not given through objects but can be executed only by the subject itself, since it is an act of its self-activity. (CPR, B130f.)

Strictly speaking, the deduction, i. e. justification of the categories (as having objective reality) only applies for what Kant calls objective representations, namely, concepts and

intuitions. In this context, he is not concerned with merely subjective representations like sensations (CPR, B376). That is, he did not intend (or at least did not demonstrate) that the categories are necessary for sensations and their qualities, presumably because they do not claim to be objective and count as knowledge in the sense that the subsumption of sensual input under concepts claims to be objective. This is in line with the contemporary notion of 'qualia' as merely subjective and ineffable qualities of experience; in a functionalist framework, qualia are not considered to be cognitively significant. Indeed, it is this very assumption that has led to famous objections against functionalism (Shoemaker 1975). Yet one might argue that acts of synthesis that do not employ concepts apply to sensations after all, such that every conscious experience, including aesthetic experience, involves some kind of synthesis (in this case performed by the imagination). Since Kant was concerned with thinking, the relevant *combinations* that he had in mind contain *logical* relations among concepts (or among judgments). The execution of acts of synthesis performed by the understanding is called thinking. The result of this process is a thought that can claim to be objectively true (In the A-Edition, Kant of course outlines various forms of synthesis performed by the imagination that do not lead to thoughts but to images). Therefore, it is somewhat strange that Hanna and Thompson align Kant's notion of spontaneity to chaotic and irregular bursts of self-generated neural activity. In Kant's theory, acts of spontaneity are expressions of logical laws, and are supposed to provide regularity, not irregularity. Since spontaneity is associated with regular synthesis, it cannot also be associated with chaos and irregularity.

Similarly, Northoff's suggestion that Kant's notion of spontaneity should be identified with intrinsic resting state activity that precedes and modulates stimulus-driven neural activity seems problematic as well: the connection between resting state activity and spontaneity associated with synthesis seems unjustified, since it has not been demonstrated that the former is involved in some kind of synthesis process. What remains a possibility is that the resting state provides a necessary background state that ensures self-reference but is independent of any specific stimulus. Then it could be seen as the neural analogue of what, in Kant's theory, is the "transcendental" self: the source of spontaneity and all activities of structuring. But this interpretation is in tension with the evidence that also associates the resting state with mindreading abilities, i. e. the ascription of mental states to others. Furthermore, since Kant leaves the specification of the transcendental "I" completely undetermined, such an analogy presupposes that we have a neural function whose role remains totally underdetermined. But this is no longer true for the resting state. While the identification of an intrinsic resting state activity is by itself an important discovery in cognitive neuroscience, we take it as an important task for future research to further clarify and specify the function of this activity in the context of perception and cognition. Yet this does not seem to illuminate Kant's account of spontaneity in any way.

As we saw above, Thompson (2007) associates Kant's notion of spontaneity with "inner purposiveness", which is not only a central notion of enactivism as developed by Varela et al. (1991), but also of Kant's philosophy of biology, namely his understanding of organisms. Therefore, the last section will be devoted to the main ideas of Thompson's "autopoietic enactivism", and to some important analogies with Kant's philosophy of organisms as self-organizing and self-producing systems.

4. Mind, life, and autopoietic enactivism

In his *Critique of Judgement*, Kant formulates a philosophy of the organism that systematically appeals to teleological explanation, and situates this kind of explanation in relation to mechanistic explanations in the natural sciences. Here, Kant conceives of organisms as 'self-organized' and 'self-producing', i. e. autopoietic systems that cannot be explained purely mechanistically, but which we have to 'make intelligible' by relying on teleological principles that are not part of natural science but *borrowed* from practical contexts. He discusses examples to demonstrate that animals exhibit a certain *form* or *organization* that, if conceived *merely* as the result of blind mechanistic causal processes, appears completely contingent. Yet, even if we have to assume that animals – as natural products – result from such mechanisms, this explanation is insufficient for our understanding:

> Since reason must be able to cognize the *necessity* in every form of a natural product if it would understand the conditions connected with its generation, the *contingency* of their form with respect to all empirical laws of nature in relation to reason is itself a ground for regarding their causality as if it were possible only through reason; but this is then the capacity for acting in accordance with ends (a will); and the object which is represented as possible only on this basis is represented as possible only as an end. (Kant 1790, 370)

Based on this feature of reason, i. e. that it aims at the intelligibility of the organizational features of organisms as being *necessitated*, Kant arrives at a conception of organisms as "natural ends". That is, Kant claims that organisms have to be conceived as natural products *and* as ends at the same time, which looks like a contradiction, since the notion of an 'end' or 'purpose' must be projected into nature but is not integral to it. Something can *either* be a product of nature or a realization of an end or purpose, but not both at the same time. So we borrow the notion of an 'end' from our practical context of producing artefacts, and introduce it for the sake of our understanding of nature and of how organisms generate.

When we produce an artefact like a watch, the *idea* of the product (in the designer) precedes and guides the chain of effective causes that eventually produce the watch. In the case of organisms, no such designer of nature can be introduced. As Kant says explicitly in the quote above, and in many other places in the *Critique of Judgement*, it is only regarded *as if* organisms were possible only through reason. Nowhere does he exclude the possibility that in fact they *are* possible through mechanistic causes. Indeed, he takes it to be impossible to prove that they *cannot* be the result of mechanistic causes. Thus, one maxim of the power of judgment

> indicates that I *should* always *reflect* on them (i. e. events in material nature, TS) *in accordance with the principle* of the mere mechanism of nature, and hence research the latter, so far as I can, because if it is not made the basis for research then there can be no proper cognition of nature. (Kant 1790, 387)

In contrast to mere mechanistic causality, organisms are supposed to exhibit a kind of "circular causality" (Thompson 2007, 62), i. e. a causal dependence which goes two ways: on the one hand, the features of the whole (organism) are determined by its parts (organs); on the other hand, the local interactions of the parts (organs) are determined by the whole (organism). Furthermore, in contrast to the watch, the parts of the organism, its organs, *produce themselves* rather than being produced by an external power, and they arrange

themselves in relation to each other. So, in contrast to the watchmaker, in the case of organisms the guiding *idea* is not to be found outside the product (the watch), but within it (the organism itself). Therefore,

> an organized being is thus not a mere machine, for that has only a *motive* power, while the organized being possesses in itself a *formative* power, and indeed one that it communicates to the matter, which does not have it (it organizes the latter). (Kant 1790, 374)

Yet, when Kant declares the intrinsic purposiveness assumed in the teleological description of the generation of organisms to be merely a regulative principle for our power of judgement, he thereby claims that we cannot *prove* that organisms indeed exhibit this formative power, since we cannot intuit it, which would be necessary in order to prove the objective reality of the notion of a natural end. But since this is due to our cognitive closure, it can neither be said that the generation of organisms is *impossible* by mere mechanisms (it may well be possible although we may not find this intelligible), nor can it be demonstrated that nature *entails* such a causality in relation to ends, a "technique of nature" (Kant 1790, 404): therefore, "the concept of that causality is a mere idea, to which one by no means undertakes to concede reality, but uses only as a guideline for reflection" (Kant 1790, 389).

In one of his latest publications, Thompson's mentor and collaborator Francisco Varela regards Kant's position as important, because he had "developed the possibility of a third way between a strong teleology and a brute materialism" (Weber & Varela 2002, 99). Yet he also considers Kant's position "unstable" and in need of revision "on the basis of modern developments of biological research and thinking". According to Weber and Varela, Kant's conception of an organism as a self-organized and self-producing being is closely analogous to the definition of an organism in Varela's own theory of "autopoiesis". On this view, biological autonomy and individuality lead to an "intrinsic teleology", to the effect that "*organisms are subjects having purposes according to values encountered in the making of their living*" (ibid., 102). This, though, is the opposite of a reductionist program, since it introduces subjectivity and value into biology.

The explanatory tool that Varela is thinking of (and which was missing in Kant's time) is a convincing empirical theory of self-organization and self-production. This is the theory of autopoiesis (on which Thompson's autopoietic enactivism is based):

> An autopoietic system – the minimal living organization – is one that continuously produces the components that specify it, while at the same time realizing it (the system) as a concrete unity in space and time, which makes the network of production of components possible. (ibid., 115)

According to Weber and Varela (and Thompson agrees), the theory of autopoiesis as a theory of living systems can help to *naturalize* Kant's original theory of organisms as self-organized and self-producing systems. The details of this account need not concern us here (cf. Thompson 2007 and Schlicht 2011 for a critical evaluation of Thompson's account). Suffice it to say that Kant, though he did not have the explanatory tools of modern biology and dynamical systems theory at his disposal, paved the way for this systematic alternative that is not reductionist and does not lead to an appraisal of strong teleology in nature, as Thomas Nagel (2013) has recently proposed.

Conclusion

In this paper, we have investigated the possible bridges between some of Kant's central claims about consciousness, mental content, and functions of mind in general, and some specific treatments of these topics in contemporary philosophy of mind and cognitive science. Since Kant did not focus on developing a thorough account of all these aspects of mind, the discussion in this paper should not be read as an exegetical exercise in a narrow sense. Nevertheless, Kant discovered many important features of the mind that are either still alive in contemporary debates, or have only recently been rediscovered. We demonstrated in the first section that Kant can be understood as defending non-conceptual representational content such that we cannot attribute to him (as has been done occasionally) either the thesis of the primacy of the propositional for all mental phenomena, or an extreme version of cognitive penetration according to which all mental phenomena are shaped by conceptual knowledge. In the second section, we outlined striking parallels between Kant's view of the difference between conscious and unconscious representations and the global workspace model of consciousness. Moreover, Kant's original view can be used as an inspiration for an alternative to the dominant versions of representationalism with respect to phenomenal consciousness. While all existing versions of the latter make a certain kind of content responsible for the difference between conscious and unconscious representations, Kant's view suggests an approach according to which a specific process of integration is responsible for this difference. According to this alternative, the relevant condition that marks the difference between conscious and unconscious representations is a process of integration that is not only popular in various philosophical accounts, but also extensively supported by cognitive neuroscience, e. g. the global workspace theory. In the third section, we investigated recent interpretations of Kant's notion of spontaneity in terms of intrinsic or self-generated brain activity and found them wanting. Neither chaotic bursts of neural activity nor the brain's resting state seem the right candidates for a neural implementation of what Kant identifies as the mind's spontaneity. Finally, the fourth section outlined a strong connection between Kant's theory of organisms and its naturalistic re-appraisal in the theory of autopoiesis. In sum, once we set aside the fact that Kant's general transcendental approach to the mind is anti-empiricist, and turn our attention to specific topics and claims regarding the mind, it turns out that there are many bridges connecting Kant with contemporary cognitive science. But just as Brook highlights important Kantian claims that have not been taken up by cognitive science, we have tried to demonstrate that other mere putative links between Kant and cognitive science remain only apparent.[4]

[4] We are very grateful for valuable comments by Andrew Brook on an earlier version of this paper. Tobias Schlicht is supported by a research grant from the Volkswagen foundation.

References

CPR Kant, I. 1781/1787. *Critique of Pure Reason*, P. Guyer/A.W. Wood. (ed./trs.) 1998. Cambridge: Cambridge UP.

Allison, H. 1996. *Idealism and Freedom. Essays on Kant's theoretical and practical philosophy*. Cambridge: Cambridge UP.

Allison, H. 2001. *Kant's Theory of Taste*. Cambridge: Cambridge UP.

Ameriks, K. 2000. *Kant's Theory of Mind. An Analysis of the Paralogisms of Pure Reason*. Oxford: Oxford UP (2nd ed.).

Baars, B. 1988. *A Cognitive Theory of Consciousness.* Cambridge: Cambridge UP.

Bayne, T./Montague, M. (eds.) 2011. *Cognitive Phenomenology*. Oxford: Oxford UP.

Bayne, T. 2010. *The Unity of Consciousness.* Oxford: Oxford UP.

Bayne, T./Chalmers, D. J. 2003. What Is the Unity of Consciousness? In: Cleeremans, A. (ed.). *The Unity of Consciousness*. *Binding, Integration and Dissociation*. Oxford: Oxford UP, 23–58.

Bermúdez, J.-L./Cahen, A. 2008. Nonconceptual Mental Content. In: Zalta, E. (ed.). *Stanford Encyclopedia of Philosophy*. URL = http://stanford.library.usyd.edu.au/entries/content-nonconceptual/.

Biro, J. 2004. Cognitive Science and David Hume's science of the mind. In: Korta, K./Larrazabal, J. M. (eds.). *Truth, Rationality, Cognition and Music*. Dordrecht: Kluwer.

Block, N. 2015. The Canberra Plan Neglects Ground. In: Horgan, T./Sabatés, M. M./Sosa, E. (eds.). *Qualia and Mental Causation in a physical world*. Cambridge: Cambridge University Press, 105–133.

Block, N. 2011. Perceptual Consciousness Overflows Cognitive Access. *Trends in Cognitive Sciences 12*, 567–575.

Boden, M. A. 2007. *Mind as Machine: A History of Cognitive Science.* Oxford: Clarendon.

Brandom, R. 2002. Reasoning and Representing. In: Chalmers, D. J. (ed.). *Philosophy of Mind*. Oxford: Oxford UP.

Brook, A. 1994. *Kant and the Mind*. Cambridge: Cambridge UP.

Brook, A. 2007. Kant and Cognitive Science. In: Brook, A. (ed.). *The Prehistory of Cognitive Science*. Basingstoke, UK: Palgrave Macmillan, 117–136.

Buckner, R. L. 2012. The Serendipitous Discovery of the Brain's Default Network. *Neuroimage 62*, 1137–1145.

Chalmers, D. J. 1996. *The Conscious Mind. In Search of a Fundamental Theory.* Oxford: Oxford UP.

Chemero, A. 2009. *Radical Embodied Cognitive Science*. Cambridge, Mass.: MIT Press.

Clark, A. 1997. *Being There: Putting Brain, Body and World Together Again*. Cambridge, Mass.: MIT Press.

Craig, A. D. 2009. How Do You Feel – Now? The Anterior Insula and Human Awareness. *Nature Reviews Neuroscience 10*, 59–70.

Crick, F./Koch, C. 1990. Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences 2*, 263–275.

Damasio, A. 1999. *The Feeling of What Happens*. New York: Harcourt.

Damasio, A. 2011. *Self Comes to Mind*. New York: Pantheon.

Dehaene, S./Changeux, J.-P./Naccache, L./Sackur, J./Sergent, C. 2006. Conscious, Preconscious, and Subliminal Processing: a Testable Taxonomy. *Trends in Cognitive Sciences 10(5)*, 204–211.

Edelman, G. M./Tononi, G. 2000. *Consciousness: How Matter Becomes Imagination.* London: Basic Books.

Engel, A. K./Fries, P./König, P./Brecht, M./Singer, W. 1999. Temporal Binding, Binocular Rivalry, and Consciousness. *Consciousness and Cognition 8*, 128–151.

Evans, G. 1982. *The Varieties of Reference*. Oxford: Oxford UP.

Fazelpour, S./Thompson, E. 2015. The Kantian Brain: Brain Dynamics from a Neurophenomenological Perspective. *Current Opinion in Neurobiology 31*, 223–229.

Fodor, J. 1983. *The Modularity of Mind*. Cambridge, Mass: MIT Press.

Fodor, J. 1975. *The Language of Thought.* Cambridge, Mass.: MIT Press.

Gennaro, R. (ed.). 2004. *Higher-Order Theories of Consciousness.* Amsterdam: John Benjamin's.

Gunther, Y. (ed.). 2003. *Essays on Nonconceptual Content*. Cambridge, Mass.: MIT Press.

Hanna, R. 2005. Kant and Nonconceptual Content. *European Journal of Philosophy 13*, 247–290.

Hanna, R./Thompson, E. 2003. Neurophenomenology and the Spontaneity of Consciousness. *Canadian Journal of Philosophy Suppl. 29*, 133–162.

Heidemann, D. H. 2012. The 'I Think' Has to Be Possible to Accompany All My Representations. Kant's Theory of Apperception and the Unconscious. In: Giordanetti, P./Rozzo, R./Sgarbi, M. (eds.). *Kant's Philosophy of the Unconscious*. New York: de Gruyter, 37–60.

Hutto, D. D./Myin, E. 2013. *Radicalizing Enactivism. Basic Minds without Content*. Cambridge, Mass.: MIT Press.

Jacobson, A. J. 2007. Doing it his way: Hume's Theory of ideas and contemporary cognitive science. In: Brook, A. (ed.). *The prehistory of cognitive science*. Vasingstoke, UK: Palgrave MacMillan, 97–116.

Kant, I. *Kants Werke*. Akademie-Textausgabe. Unveränderter photomechanischer Abdruck von Kants gesammelten Schriften, Königlich Preußische Akademie der Wissenschaften (ed.) 1968. Berlin: de Gruyter.

Kant, I. 1790. *Critique of the Power of Judgement,* Guyer (ed.) 2002. Cambridge: Cambridge UP. (Original Kant, I. 1790. *Kritik der Urteilskraft,* Berlin.)

Kant, I. 1798. *Anthropology from a Pragmatic Point of View*, Louden/Kuehn (ed./trs.) 2006. Cambridge: Cambridge UP. (Original: Kant, I. 1798. *Anthropologie in pragmatischer Hinsicht*, Königsberg.)

Keysers, C./Gazzola, V. 2007. Integrating Simulation and Theory of Mind: From Self to Social Cognition. *Trends in Cognitive Sciences 11,* 194–196.

Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton UP.

Kitcher, P. 1990. *Kant's Transcendental Psychology*. Oxford: Oxford UP.

Kriegel, U. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford and New York: Oxford UP.

Lamme, V. 2006. Towards a True Neural Stance on Consciousness. *Trends in Cognitive Sciences 10(11),* 494–501.

LeDoux, J. 1999. Psychoanalytic Theory: Clues from the Brain. *Neuropsychoanalysis 1*, 44–49.

MacPherson, F. 2012. Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism. *Philosophy and Phenomenological Research 84(1)*, 24–62.

Levine, J. 2001. *Purple Haze.* Cambridge, Mass.: MIT Press.

Lupyan, G. 2012. Linguistically Modulated Perception and Cognition: The Label-feedback Hypothesis. *Frontiers in Psychology 3*, 54.

Marchi, F./Newen, A. 2015. Cognitive Penetrability and Emotion Recognition in Human Facial Expressions. *Frontiers in Psychology 6(828)*, DOI=10.3389/fpsyg.2015.00828.

Marcel, A. J. 1983. Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition. *Cognitive Psychology 15*, 197–237.

McDowell, J. 1994. *Mind and World.* Cambridge, Mass: MIT Press.

Meerbote, R. 1989. Kant's Functionalism. In: Smith, J.C. (ed.). *Historical Foundations of Cognitive Science*. Dordrecht: Kluwer.

Mole, C. forthcoming. Attention and Cognitive Penetration. In: Zembeikis, J./Raftopoulos, T. (eds.). *Cognitive Penetration*. Oxford: Oxford UP.

Merker, B. 2007. Consciousness Without a Cerebral Cortex: A Challenge for Neuroscience and Medicine. *Behavioral and Brain Sciences 30*, 63–134.

Metzinger, T. (ed.) 2000. *Neural Correlates of Consciousness.* Cambridge, Mass.: MIT Press.
Metzinger, T. 1995. Faster than Thought. In: Metzinger, T. (ed.). *Conscious Experience.* Thorverton: Imprint Academic, 425–464.
Moruzzi, G./Magoun, H. 1949. Brain Stem Reticular Formation and Activation of the EEG. *Electroencephalography and Clinical Neurophysiology 1*, 455–473.
Nagel, T. 1974. What Is It Like to Be a Bat? *Philosophical Review 83*, 435–456.
Nagel., T. 2013. *Mind and Cosmos.* Oxford: Oxford UP.
Nakano, H. 2011. Selbstaffektion in der transzendentalen Deduktion. *Kant-Studien 102(2)*, 213–231.
Newen, A./Vogeley, K. 2008. Grundlegende Paradigmen in der Philosophie des Geistes: Die Grenzen der Transzendentalphilosophie und die Zukunft des Naturalismus. In: Spät, P. (ed.). *Zur Zukunft der Philosophie des Geistes*. Paderborn: mentis, 93–124.
Newen, A. 2003. David Hume. Die Transparenz des Geistes sowie das Ich als Bündel und Einheit von Perzeptionen. In: Newen, A./Meixner, U. (eds.). *Seele, Denken Bewusstsein. Zur Geschichte der Philosophie des Geistes.* Berlin: de Gruyter, 232–284.
Northoff, G./Quin, P./Nakao, T. 2010. Rest-stimulus Interaction in the Brain: A Review. *Trends in Neurosciences 33*, 277–284.
Northoff, G. 2012. Immanuel Kant's Brain and the Brain's Resting State. *Trends in Cognitive Sciences 16*, 356–359.
Northoff, G. 2013. What the Brain's Intrinsic Activity Can Tell Us About Consciousness? A Tridimensional View. *Neuroscience and Biobehavioral Reviews 37*, 726–738.
Pippin, R. 1987. Kant on the Spontaneity of Mind. *Canadian Journal of Philosophy 17(2)*, 449–475.
Pylyshyn, Z. W. 1999. Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception. *Behavioral and Brain Sciences 22*, 341–423.
Raffman, D. 1995. On the Persistence of Phenomenology. In: Metzinger, T. (ed.). *Conscious Experience.* Paderborn: Schöningh, 293–308.
Rosenthal, D. M. 2004. *Consciousness and Mind*. Oxford: Oxford UP.
Sadaghiani, S./Scheeringa, R./Lehongre, K./Morillon, B./Giraud, A. L./Kleinschmidt, A. 2010. Intrinsic Connectivity Networks, Alpha Oscillations, and Tonic Alertness: A Simultaneous Electroencephalography/Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience 30(30)*, 10243–10250.
Schlicht, T. 2007. *Erkenntnistheoretischer Dualismus. Das Problem der Erklärungslücke in Geist-Gehirn-Theorien*. Paderborn: mentis.
Schlicht, T. (ed.). 2011. Consciousness and Life. Special Issue of *Journal of Consciousness Studies 18(5,6)*.
Schlicht, T. 2012. Phenomenal Consciousness, Attention, and Accessibility. *Phenomenology and the Cognitive Sciences 11(3)*, 309–334.
Searle, J. 2005. *Mind. A brief introduction.* Oxford: Oxford UP.
Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.
Sellars, W. 1974. '… this I or he or it (the thing) which thinks …'. Immanuel Kant, Critique of Pure Reason (A346/B404). In: Sellars, W. (ed.). *Essays in Philosophy and its History*. Dordrecht: Kluwer, 341–362.
Shewmon, D./Holmse, D./Byrne, P. 1999. Consciousness in Congenitally Decorticate Children: Developmental Vegetative State as a Self-fulfilling Prophecy. *Developmental Medicine and Child Neurology 41*, 364–374.
Shoemaker, S. 1975. Functionalism and Qualia. *Philosophical Studies 27*, 291–315.
Siegel, S. 2012. Cognitive Penetrability and Perceptual Justification. *Noûs*, 46(2), 201–222.
Singer, W. 2013. Cortical Dynamics Revisited. *Trends in Cognitive Sciences 17*, 616–626.
Solms, M. 2014. The Conscious Id. *Neuropsychoanalysis 15(1)*, 5–19.
Stokes, D. 2013. Cognitive Penetrability of Perception. *Philosophy Compass 8(7)*, 646–663.
Sturma, D. 1985. *Kant über Selbstbewusstsein*. Hildesheim: Olms.

Thompson, E. 2007. *Mind in Life. Biology, Phenomenology and the Sciences of the Mind*. Cambridge, Mass: Harvard UP.

Tononi, G./Koch, C. 2008. The Neural Correlates of Consciousness. An Update. *Annual Review of the N.Y. Acad. of Sciences 1124*, 239–261.

Tye, M. 1995. *Ten Problems of Consciousness.* Cambridge, Mass.: MIT Press.

Van Gulick, R. 2004. Higher-order Global States. In: Gennaro, R. (ed.). *Higher-order Theories of Consciousness*. Amsterdam: John Benjamin's, 67–92.

Van Gulick, R. 2006. Mirror Mirror – Is that all? In: Kriegel, U./Williford, K. (eds.). *Self-representational Approaches to Consciousness*. Cambridge, Mass: MIT Press, 11–40.

Varela, F./Thompson, E./Rosch, E. 1991. *The Embodied Mind*. Cambridge, Mass.: MIT Press.

Vetter, P./Newen, A. 2014. Varieties of Cognitive Penetration in Visual Perception. *Consciousness & Cognition 27*, 62–75, doi:10.1016/j.concog.2014.04.007.

Vogeley, K./Bussfeld, P./Newen, A./Herrmann, S./Happé, F./Falkai, P./Maier, W./Shah, N. J./Fink, G. R./Zilles, K. 2001. Mind Reading: Neural Mechanisms of Theory of Mind and Self-perspective. *Neuroimage 14*, 170–181.

Vosgerau, G./Schlicht, T./Newen, A. 2008. Orthogonality of Phenomenality and Content. *American Philosophical Quarterly 45*, 329–348.

Weber, A./Varela, F. 2002. Life after Kant. Natural Purposes and the Autopoietic Foundations of Biological Individuality. *Phenomenology and the Cognitive Sciences 1*, 97–125.

Weiskrantz, L. 1986. *Blindsight*. Cambridge, Mass.: MIT Press.