Consciousness and Cognition 29 (2014) 248–258

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Review

From commonsense to science, and back: The use of cognitive concepts in neuroscience

Jolien C. Francken [a,*], Marc Slors [b]

[a] Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, Netherlands
[b] Department of Philosophy, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, Netherlands

A R T I C L E I N F O

Article history:
Received 28 February 2014
Available online 6 October 2014

Keywords:
Concepts
Ontology
Philosophy of mind
Interpretivism
Mental realism
Cognitive neuroscience
Attention
Consciousness
Memory
Cognitive control

A B S T R A C T

Commonsense cognitive concepts (CCCs) are the concepts used in daily life to explain, predict and interpret behaviour. CCCs are also used to convey neuroscientific results, not only to wider audiences but also to the scientific inner circle. We show that translations from CCCs to brain activity, and from brain data to CCCs are made in implicit, loose and unsystematic ways. This results in hard to connect data as well as possibly unwarranted extrapolations. We argue that the cause of these problems is a covert adherence to a position known in philosophy of mind as 'mental realism'. The most fruitful way forward to a clearer and more systematic employment of CCCs in cognitive neuroscience, we argue, is to explicitly adopt interpretivism as an alternative for mental realism. An interpretative stance will help to avoid conceptual confusion in cognitive science and implies caution when it comes to big conclusions about CCCs.

Contents

* Corresponding author. Fax: +31 24 36 10652.
E-mail addresses: j.francken@donders.ru.nl (J.C. Francken), m.slors@ftr.ru.nl (M. Slors).

http://dx.doi.org/10.1016/j.concog.2014.08.019
1053-8100/

1. Introduction

Despite seven decades of hard work on rabbits, rats, mice, gerbils, guinea pigs, sheep, cats, dogs, Old World monkeys, chimpanzees, and humans by outstanding colleagues, to date, there is still no agreed term that would unequivocally describe behavioural correlate(s) of hippocampal theta rhythms. [...] [A] sober conclusion is that our behavioural-cognitive terms are simply working hypothetical constructs that do not necessarily correspond to any particular brain mechanism (Buzsáki, 2006).

Cognitive neuroscience deals with data such as BOLD signals,[1] electrophysiological data and reaction times. Such data may be of direct interest to neuroscientists, but their relevance stems mainly from the fact that they tell us something about the human *mind*. This is specifically true from the perspective of the lay public, which is generally enthusiastic about cognitive neuroscience, but which only wants to hear about specific neural activity because this is considered to tell them something about cognitive functions they are familiar with, such as thinking, loving and choosing. Hence, cognitive concepts are used to convey neuroscientific results, not only to wider audiences but also to the scientific inner circle. Scientific publications on the possible neural correlates of e.g. consciousness, romantic love, hate, personality, decision making, error awareness, etc. are common. Expressing what makes cognitive neuroscience exciting – the fact that it tells us something about *us*; about who we are and what makes us tick – requires the use of commonsense cognitive concepts.

This fact makes it imperative that we have a clear and systematic view on how we get from commonsense cognitive concepts to scientific experiments and from hard scientific data to commonsense concepts. In this paper we will argue that so far these issues are at best addressed in an implicit manner in cognitive neuroscience. Transitions from commonsense psychology to hard science and back are often made without explicit ideas on how mind and brain relate. We argue that this results in incompatible operationalisations of 'the same' cognitive concept and consequently hard to connect data and possibly unwarranted extrapolations.

To give an initial sense of one kind of problem we have in mind, here is an example of a single cognitive concept that is used by different researchers to refer to different cognitive processes, measured by different tasks. The concept of 'working memory' (Poldrack et al., 2011) is interpreted by Goldman-Rakic as 'holding information online' when working with non-human primates using tasks such as the oculomotor delayed response task (Goldman-Rakic, 1995). Baddeley uses the concept as 'manipulating information held in memory' by humans, using e.g. the letter–number sequencing task (Baddeley, 1992). Olton et al. use the concept as a rough equivalent of episodic memory, that is as memory for temporally varying aspects of a task, when measuring rodents using a radial arm maze task with varied food locations (Olton, Becker, & Handelmann, 2011). As a result of this ambiguous use of the concept 'working memory', it is difficult to combine insights from different studies into convergent knowledge about the neuroscientific basis of working memory.

Disentanglement of these kinds of difficulties starts with adopting an explicit view on how commonsense cognitive concepts relate to operationalisations and tasks and to the neuronal processes that allow us to execute these tasks. We need a theory of the mind–brain relation that fits the actual practices of cognitive neuroscience. In this paper we will argue that most classical positions in the philosophy of mind are not up to that job and tentatively propose what we consider to be the most useful view on the relation between mind and brain.

This paper is organized as follows. In Section 2, we show how cognitive commonsense concepts are used in cognitive neuroscience and why this use is problematic. Section 3 describes the development of 'cognitive ontologies' as a means of dealing with these issues. The success of this strategy depends on a clear view on the root of the problem. In Section 4 we identify this as a covert adherence to mental realism. Based on the discussion in the previous three sections, we argue that the most fruitful way forward to a clearer and more systematic employment of commonsense cognitive concepts in cognitive neuroscience depends on the acceptance of an interpretivist view on the relation between mind and brain.

2. Commonsense cognitive concepts in cognitive neuroscience: problematic translations at multiple levels

Commonsense cognitive concepts (CCCs), also referred to as 'folk-psychological concepts', are the concepts we use in daily life to explain, predict and interpret behaviour. 'Thought', 'intention', 'fear', 'joy', 'disgust', 'consciousness', 'attention', 'desire', 'belief', 'wish', etc. are cases in point. How are these CCCs used in cognitive neuroscience? And how are neuroscientific data – such as BOLD signals, electrophysiology data, reaction times – translated in CCC-terminology? How are questions about the human mind, which are very often cast in CCC terms, operationalised in experiments? In this section we will argue that such transitions are often made in a loose, unsystematic and implicit manner.

Here it may be objected beforehand that much cognitive neuroscience is not concerned with rough and vague commonsense concepts but rather with scientifically 'cleaned-up' versions of these. We agree. We shall refer to such concepts as SCCs, scientific cognitive concepts. Scientific cognitive concepts are formal, scientific versions of common sense cognitive concepts. Often, the particular concepts have the same name (e.g. 'memory') but the SCCs usually differ from CCCs e.g. in their sectioning in sub-concepts (see Fig. 1). Thus, the SCC 'memory' differs from the CCC 'memory' in that the former is composed of more specifically delineated SCCs such as 'working memory' and 'long-term memory', to name a few. The scientific status of the SCC 'memory' results from the precision of these sub-concepts. Ideally, SCCs are formalized versions of CCCs with clearer definitions that are shared by the entire scientific community. We shall argue that though SCCs are certainly more

[1] The fMRI (functional magnetic resonance imaging) signal is usually referred to as the blood oxygen level-dependent (BOLD) signal since fMRI relies on changes in the level of oxygen in the human brain induced by alterations in blood flow.

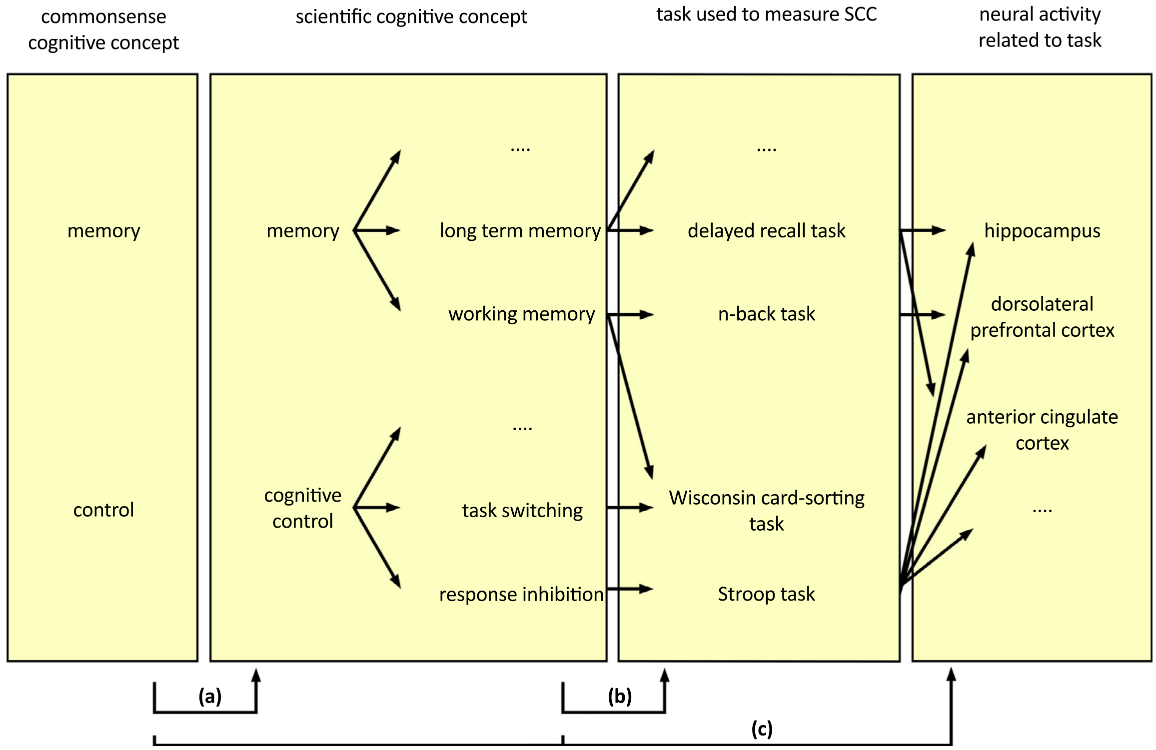

Fig. 1. The different levels between commonsense cognitive concepts (CCCs) and neural data: scientific cognitive (sub-)concepts (SCCs) and task operationalisations. A. From commonsense cognitive concepts ‘memory’ and ‘control’ the respective scientific cognitive concepts ‘memory’ and ‘cognitive control’ are constructed. These scientific concepts are sectioned in sub-concepts. See also Section 2.1. B. Particular experimental tasks are used to operationalise specific scientific cognitive concepts. For instance, the ‘n-back task’ and ‘Wisconsin card-sorting task’ are both used to investigate ‘working memory’. See also Section 2.2. C. CCCs and SCCs are sometimes directly linked to neural mechanisms (e.g., when measured with fMRI, BOLD activity in specific brain regions). See also Section 2.3.

scientifically respectable, they will not help to allay the problems that stem from the lack of a systematic view on the relation between CCCs and brain data.

Fig. 1 shows the four layers or strands that take us from CCC to brain activity: (1) CCCs are cleaned up as (2) SCCs, which are operationalised in terms of (3) specific tasks; execution of these tasks is correlated with (4) neural activity (e.g., when measured with fMRI, BOLD activity in specific brain regions). The examples of the CCCs ‘memory’ and ‘control’ show that there are multiple lines from CCCs to activity in specific brain regions. In fact there are so many bifurcations in this scheme that the translations from CCC/SCC to brain activity and from brain data to SCC or CCC are highly indirect and prone to various interpretations. This leads to a serious lack of clarity when experimental findings are expressed in terms of CCCs.

In the following subsections, we will discuss examples at three different levels in our scheme illustrating that the translations are made in loose, implicit and unsystematic ways.

2.1. From CCCs to SCCs

One problem is that while specific SCCs or sub-concepts of SCCs are the target of certain studies, these studies are presented as being about the CCCs from which the relevant SCC is derived (**see Fig. 1, arrow a**). For instance, in two arbitrarily chosen cognitive neuroscience papers, ‘attention’ is defined either as ‘whether a stimulus is behaviourally relevant’ (Kok, Rahnev, Jehee, Lau, & de Lange, 2012) or as ‘where/when a stimulus is likely to occur’ (Van Ede, de Lange, Jensen, & Maris, 2011). In these cases, it is not explicitly recognized that these SCCs capture only part of the CCC ‘attention’. Neither of these SCCs capture e.g. ‘focus’ or ‘direction of thoughts at some specific task’, which are also important aspects of attention as a CCC.

2.2. From SCCs to task operationalisations

SCCs such as ‘working memory’ are vaguely defined: different experiments highlight different interpretations and/or aspects of it. This scattering of broad, vaguely defined SCCs into an array of more narrowly defined sub-concepts is driven by the need to operationalise SCCs in such terms as to yield unambiguous data (**see Fig. 1, arrow b**). Problems occur when cognitive neuroscientists *define* their SCCs in terms of their operationalisations in an experimental task, implying that the tasks measure a specific SCC in its entirety. For example, the Sternberg item recognition task is often referred to as the 'Sternberg working memory task' (Poldrack et al., 2011). In this case, sub-concepts of working memory related to the Sternberg task are not clearly identified as such and hence confused for the entire SCC of working memory. But when it is not explicitly recognized that it is not the *entire* SCC but only a sub-concept that is being investigated, the result is an array of hard to connect data.

A further problem with connecting SCCs to operationalisations is that one particular task can be used to measure the underlying mechanisms of various different cognitive concepts. This is illustrated in Fig. 1: the Wisconsin card-sorting task is used to measure both 'working memory' and 'task-switching' (and several other cognitive functions) (Berg, 1948; Keefe, 1995). A similar problem occurs with the 'Stroop task' (Bilder et al., 2009). From the Stroop task, two different outcome measures can be obtained: on the one hand, the time to complete colour-word interference relative to the control condition, and on the other hand, the time to complete the colour-naming condition. The former is a measure for 'response inhibition'; the latter is an indicator for colour naming skill. As a consequence, the Stroop task cannot simply be taken as a measure for response inhibition.

2.3. From CCCs/SCCs to neural mechanisms

Another serious problem occurs when CCCs or SCCs are used to designate not their operationalisations, but the neural activation associated with performing the tasks set by these operationalisations (**see Fig. 1, arrow c**). The problem here is that while CCCs or SCCs provide the *explanandum* (i.e. that what needs to be explained) for cognitive neuroscientific research, they are sometimes used to refer to the *explanans* (i.e. the neural phenomena that are thought to do the explaining). This results in circular explanations.

In neuroimaging papers, it is common to find arguments such as 'we found activation in the striatum, showing that our condition A produced more reward than condition B' (see for examples Poldrack, 2006). This is especially the case for unexpected neural activations. Reasoning backwards from the presence of brain activation to the engagement of a particular cognitive function is called 'reverse inference' (Poldrack, 2006). Reverse inference might sometimes be a useful strategy to infer cognitive processes, e.g., for sensory or motor regions, which are known to be very selectively activated by respectively sensory or motor events. Yet for many brain regions, it is very unclear in which cognitive processes they are involved (see Fig. 1). Anderson and Pessoa (2011) quantified the cognitive diversity of individual brain regions and found indeed that most regions contribute to tasks across multiple cognitive domains. Importantly, this variability may result from the fact that individual brain regions are just not selective for specific cognitive functions but also as a consequence of conceptual unclarity and unsystematicity of our mental taxonomy.

Our next example illustrates another reason why it is problematic to omit the operationalisation step and turn directly from SCC to neural mechanisms. Price and Friston (2005) describe how different investigators assign different SCCs to the same area (the left posterior lateral fusiform area), as a consequence of the use of different tasks. Studies of reading call this area the 'visual word form area' while people studying object recognition find this area to be sensitive to the visual attributes of animals. Still others refer to this area as the 'lateral occipital tactile-visual region'. Price and Friston argue that we should not ask, "What are the neural correlates of a cognitive operation?" but rather "What is the function associated with activation of a brain region?" because, as the example shows, task-specific SCCs do not predict activation in the other tasks.

2.4. Taking stock

The above examples illustrate types of problems that occur in countless other instances. Summarizing these sets of problems, we can conclude that CCCs are translated to cognitive neuroscientific experiments and neuroscientific data are translated in CCC terms in loose, implicit and unsystematic ways. This has several problematic consequences. It produces confusion because there is no systematic relationship between CCCs and SCCs, between SCCs and task operationalisations, and as a result between SCCs and ultimately CCCs and neural mechanisms.

The problems that we describe arise when cognitive neuroscience applies concepts taken from an everyday vocabulary. However, the same type of problem occurs when cognitive neuroscientists collaborate with colleagues from different disciplines, because this requires translations across species or across different fields. This may result in problems such as the following. Fisher explains how a misconception about the way in which genes relate to cognition results in erroneous conclusions (Fisher, 2006). The misconception is that there exist straightforward linear relationships between specific genes and cognitive functions. Researchers outside genetics often forget that "genes do not specify behaviours or cognitive processes; they make regulatory factors, signalling molecules, receptors, enzymes, and so on, that interact in highly complex networks, modulated by environmental influences, in order to build and maintain the brain" (Fisher, 2006, p.270). Geneticists in part contribute to the misunderstanding by describing their results in simplified ways: "It is easier to refer to the "gene for cystic fibrosis" than "the chloride channel gene which when it is mutated gives rise to cystic fibrosis"." (Fisher, 2006, p. 291). Ultimately, assumptions of simple linear relations between genes and cognitive functions or behaviour "have impeded progress in the field, and fuel hypotheses that must ultimately be untenable" (Fisher, 2006, p. 279).

Cognitive neuroscience is an inherently interdisciplinary scientific discipline (Abi-Rached & Rose, 2010; Bilder et al., 2009; Fisher, 2006). The current conceptual problems are a major obstacle for collaborative efforts because they can result in hard to connect scientific data, both within and between disciplines, as well as unwarranted extrapolations from experiments to the wider public.

3. The development of a cognitive ontology

One important way to overcome many of the problems we describe in Section 2 is the use of formal ontologies (Bilder et al., 2009; Price & Friston, 2005). The term 'ontology' stems from philosophy where it deals with everything that exists. We rather use the term here in an information sciences sense: an ontology formally represents knowledge as a hierarchy of concepts within a domain, using a shared vocabulary to denote the types, properties and interrelationships of those concepts.

Thus, the use of ontologies gives us the opportunity to properly structure the cognitive concepts that we currently use in an unsystematic way. This will advance communication and synthesis of results within cognitive neuroscience, between cognitive neuroscience and other disciplines such as genetics and psychiatry, and the translation of scientific findings to the wider public.

The fields of genetics and biology have already benefitted from the creation of a shared ontology. These areas suffered from wide variations in terminology leading to ineffective searches for relevant information. The 'Gene Ontology project' is a collaborative effort to address the need for consistent descriptions of gene products in different databases containing information about three model organisms (worms, flies, and mice). It has developed structured controlled vocabularies that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner (Ashburner et al., 2000).

Since there is a similar need for conceptual consistency in the cognitive sciences, Poldrack and colleagues have proposed the 'Cognitive Atlas' (Poldrack et al., 2011). In this database, SCCs can be systematically related to other concepts, to the tasks that are used to assess them, and to the neural mechanisms underlying them. The things that exist in the ontology are called *instances* (e.g., the concept 'memory', or a specific experimental paradigm) and these are explicitly defined. Importantly, the *relations* between the instances are formalized. Examples of relations are 'is a kind of' (e.g., 'declarative memory' is a kind of 'memory') and 'is a test type of'. As we have shown in Section 2, this is important because the relations between for instance 'memory' and 'long-term memory' are currently not often explicit (see Fig. 1). In this way, the cognitive sciences community might start to solve the conceptual confusion.

But how should the scientific community decide on the definitions and relations? One option is to start from a top-down or theoretical perspective, e.g., using existing models from cognitive psychology, by discussing the controversies and disagreements and subsequently adding these to the database. Another, more bottom-up strategy, is to look for patterns in the current use of the SCCs. For example, Sabb and colleagues ran a search in PubMed to look for concepts that are associated with cognitive control (Sabb et al., 2008). Subsequently, again by a PubMed search, they determined which specific cognitive task measures were used most frequently to experimentally investigate those concepts. Interestingly, when a particular network of concepts and tasks is agreed upon, this hypothesis can be empirically tested. Lenartowicz et al. used the results of Sabb as a starting point to examine whether the proposed component processes of cognitive control could be mapped to independent neural systems by using classifier analyses of brain imaging data (Lenartowicz, Kalar, Congdon, & Poldrack, 2010) (see also Poldrack, Halchenko, & Hanson, 2009; Price & Friston, 2005). However, while this might give promising results for relatively coarse grained cognitive concepts, when turning to more specific cognitive concepts (i.e., most of the SCCs), we believe it will become more challenging to find unequivocal neural correlates.

A third option is to start the ontology-building project by focusing on the operationalisations before turning to the SCCs (see also Bilder et al., 2009). The main reason for this is the fact that definitions at the task level are more concrete, stable and explicit while definitions at the cognitive concept level are often absent, implicit or imprecise. For instance, when one is interested in studies about 'cognitive control', instead of using the cognitive concept as a search term in PubMed, one could search for specific tasks that are assumed to measure this cognitive function (e.g., the Stroop task). Only after assessing whether the behavioural and neural outcome measures that are obtained with this task produce similar results, one could turn to the second step. This step is to relate the Stroop task to an SCC. It is essential to check whether the task is commonly associated with just one, or with multiple cognitive constructs. Since in the latter case, the interpretation of the data would be ambiguous. Of course, more than one task can be associated with an SCC (which is the case in the 'cognitive control' example). In that case, one could repeat this procedure and try to see which aspects of the operationalisations overlap, and which do not. In sum, when starting ontology development the emphasis would be on comparisons and definitions at the task level, instead of the concept level.

These three strategies are different and not always compatible: for instance, if a given task is associated with more than one SCC, we may need to redefine SCCs so as to avoid ambiguity, and that may interfere with the idea to look for current uses of SCCs. It is important to solve such issues, for we believe that in view of the problems outlined in Section 2, ontology development is a crucial factor for the future success and progress of cognitive neuroscience. A cognitive ontology would make the definitions of concepts and tasks more explicit, resulting in facilitated communication and the potential for validation of findings by meta-analyses. For this to be possible, however, we need a clear idea of what underlies the ambiguities and problems sketched in Section 2. In the following we shall argue that the root problem here is an implicit adherence to what is known in philosophy as 'mental realism'.

4. How implicit realism results in problematic translations

4.1. Natural kinds vs. human kinds

As the previous sections witness, the use of CCCs in cognitive neuroscience is abundant. However, whereas brain states and neural processes may count as natural kinds, there are good reasons to deny CCCs and SCCs a similar status. For instance, Danziger points to the fact that changes to CCCs an SCCs do not follow from experimental results (Danziger, 1997). In fact CCCs precede these experiments by several centuries if not millennia:

Do (...) categories like cognition, emotion, learning, (...) intelligence, etc. represent natural kinds? Are we the people who happen to have hit on a nomological net that genuinely reflects the natural, the objective, divisions among classes of psychological events? Perhaps. But if we are, it is not because of our superior methods of empirical investigation. For the categories in question were not invented as a consequence of empirical investigation—they were there before anyone used them to identify the objects of empirical studies (Danziger, 1997, p. 5).

CCCs are devised primarily to function in the praxis of everyday life. Their explanatory and predictive potential is usually put in the service of facilitating everyday social interaction (cf. Cromwell & Panksepp, 2011; Uithol, Burnston, & Haselager, 2014). This explanatory and predictive use only maps *partially* onto the kind of explanations offered by cognitive neuroscience (Uithol et al., 2014). Thus, many have claimed that it is reasonable to consider CCCs 'human kinds' rather than natural kinds (e.g. Danziger, 1997; see also Hacking, 1986). In other words, CCCs need not be 'real' in any neuroscientific sense. The point here is that the use we make of CCCs in daily praxis endows them with characteristics that are not shared by neural explanations of behaviour. Let us mention three of these.

First of all, CCCs do not only explain actions causally, they often also rationalize behaviour. For instance, by claiming that I want to quench my thirst and believe there is a cold beer in the fridge, I not only provide the causal antecedents behind my getting up and walking to the kitchen, I also explain why and in what sense it is rational for me to do so. This feature of CCC explanations ties in with another characteristic: CCCs function in describing our reasons for action. By providing reasons for action, I imply or presuppose many other commitments. For instance, my thirst and my belief that there is a beer in the fridge explains my action only when I know what a fridge is, when I believe that it is in the kitchen, when I believe the kitchen is behind this door, when I believe it is safe to enter the kitchen, do not believe beer is poisonous, etc. Philosophers refer to this feature as the holism of the mental (Davidson, 1980; Dennett, 1987). Neither the rationalizing quality of reason explanation, nor the holism of the mental have clear parallels in neuroscientific explanations of behaviour.

A second feature of the explanation of behaviour by means of CCC concepts is that some such concepts have what philosophers call 'wide-content'. Beliefs that I hold and that may explain my behaviour may have contents that are not solely determined by what goes on in my head. Suppose I pay a large sum of money for a pen because I believe it is made of gold. As a matter of fact I am not an expert on gold. The meaning of the concept of 'gold' is something that I only vaguely entertain, while I rely on the existence of experts who do have a narrowly circumscribed concept of gold. The same may hold for my knowledge of the exact economic value of gold. Again I rely on experts. Hence, my belief that this pen is made of gold, which does explain my behaviour, is in an important sense not merely constituted by what goes on in my head but also by the (linguistic) community I live in (Putnam, 1988). Again, such wide content has no direct parallel in neuroscience.

Thirdly, there is considerable support for the notion that there is cultural (Lillard, 1998; Vinden, 2002) and historical (Macdonald, 2003, 2007; Uttal, 2001) variation in CCCs and also in SCCs. This variation in 'mental taxonomies' in no way corresponds with the subtle cultural differences that can be found at the level of neural processing (see e.g. Han & Northoff, 2008). Apparently the pie of the mind can be cut into pieces in various ways. This considerably undermines the naïve idea that CCC explanations can easily and straightforwardly be mapped onto neural explanations.

All this does not preclude the idea that explanations of behaviour in terms of CCCs do overlap with – and hence will turn out to coincide with – neural explanations of behaviour. What it does show, however, is that this overlap is only partial since CCC explanations are rationalizing explanations, sometimes involving wide contents, that allow for cultural variation that is not matched by concurring neural variation. CCCs as human kinds cannot be naïvely equated with brain states, brain areas or brain processes, which are natural kinds.

4.2. CCCs are assumed to be natural kinds as a result of implicit realism

The problems with applying CCCs in cognitive neuroscience that we have outlined in Section 2 arise from assuming one-on-one mappings between CCCs and brain areas or cognitive processes. For instance, if a researcher supposes there is a 'neural correlate' of the concept 'attention', he will not necessarily carefully assess all the different translational steps between the concept of attention and the neural data he will collect. The translations between the different levels – from CCCs to SCCs, to neural mechanisms and back – are made without explicit ideas on the relation between commonsense cognitive concepts and neuronal processes, in other words, the relation between mind and brain. In this subsection we will very briefly discuss the philosophical mainstream options about the mind–brain relation in order to show that the problematic assumptions about the relation between CCCs and neural processes stem from an implicit adherence to realism about CCCs. CCCs are treated as natural kinds, on a par with neural kinds.

Thoughts, feelings, intentions, etc. explain behaviour. We often cite a person's thoughts or intentions in order to provide the reasons (in a colloquial, non-technical sense) for why someone did what she did. Alternatively we can hypothesize about someone's reasons for a given action or we can predict an upcoming action if we know what that person thinks or feels. But such uses ultimately depend on there being an explanatory relation between CCCs and behaviour. In everyday discourse as well as in philosophy (at least since Davidson, 1963) this explanatory relation is generally thought to be a causal relation; thoughts, beliefs, desires, feelings, etc. are said to cause our actions.

The same holds for brain states: brain states also cause actions. Given that we know for certain that brain states cause our actions, is it not redundant to hold that the referents of CCCs cause actions too? Not necessarily. This question is debated in philosophy under the heading of 'mental causation' (Campbell, 2003; Heil & Mele, 1993).

The classical option in this debate is to hold that beliefs, desires, etc. *are* brain states – hence, the causal efficacy of mental states *is* the causal efficacy of brain states.[2] This class of mind–brain theories is called *realist* theories. The various theories that fall under this class – type-identity theories, token-identity theories and various forms of functionalism – hold that CCCs are as real as the brain states they are identical with, or on which they 'supervene' (as it is called in philosopher's jargon). Although many philosophers and scientists think that beliefs and desires are brain states, not everyone is convinced. Some hold that while both neuroscience and our CCC idiom aspire to provide causal explanations for action, neuroscience simply trumps our older folk-psychological concepts. According to this so-called eliminativism (Churchland, 1981), neuroscientific concepts will slowly replace CCCs.

Even though especially the first class of theories captures the majority of positions in contemporary philosophy of mind, we believe that the discussion in Section 2 shows that neither realism, nor eliminativism are feasible positions when it comes to elucidating the role of CCCs in the practice of neuroscientific research.

We can be brief about eliminativism. When eliminativism is correct, we should see a slow replacement of CCCs by neuroscientific explanations. The fact that CCCs and the SCCs that are derived from them are widely used both to setup neuroscientific experiments and to interpret and convey the results of those experiments shows that such replacement is simply not happening. Whether or not eliminativism is to be preferred from a theoretical standpoint, it simply does not capture the reality of the practice of neuroscience.

Realism, on the other hand, seems widely represented – albeit largely implicitly – in neuroscience. It is precisely this implicit realism that is at the root of the majority of the problems discussed in the previous section. Realism suggests that ideally there should be a one-on-one mapping of CCCs or SCCs on brain processes, however complex these may turn out to be. When CCCs or SCCs are brain processes, then it seems that each instantiation of a CCC/SCC can be identified with a (set of) brain process(es), at least in theory. Section 2 shows that in the practice of cognitive neuroscience the idea of such clear-cut mapping is highly problematic. Yet these problems are not usually acknowledged. If not, then when CCCs/SCCs provide the *explanandum* for neuroscience, and when we have (purportedly) identified the brain processes with which they are identical, then it becomes natural to refer to these processes – which are the *explanans* – whenever we want to refer to the CCC or SCC at issue. Hence, possible conflation of explanandum and explanans, as in the case of reverse inference, looms large.

Thus, even though mental realism is widespread in neuroscience, we think that in order to deal effectively with the problems described in the previous section, it should be dropped.

4.3. The solution: Interpretivism

The position on the mind–brain relation that would best suit the practice of cognitive neuroscience, then, must hold a middle ground in between realism and eliminativism. Such a position is available in many forms under the heading of 'interpretivism' (Davidson, 1980; Dennett, 1987; Mölder, 2009). One such position that comes very close to what we wish to advocate is Daniel Dennett's Intentional Stance Theory (Dennett, 1971, 1978, 1987, 1991). According to Dennett, CCCs and SCCs are concepts that track behavioural patterns.[3] They are ascribed to systems such as human beings in order to predict and explain behaviour. If such predictions and explanations prove to be fruitful, this means that the relevant CCCs/SCCs *really* apply, i.e. that people have the ascribed belief, intention, emotion, etc. But this is not because they refer to discrete neural processes that causally explain the predicted behaviour. The fruitful applicability of CCCs/SCCs to humans suggests that our brains harbour the required mechanisms that causally explain the relevant behavioural patterns. Identifying the CCC with these neural processes is, in Dennett's view, "a gratuitous bit of misplaced concreteness." (Dennett, 1987, p. 51). For beliefs, desires, intentions, etc. are states of whole persons (or other systems), not *parts* of them. Indeed, this is how these concepts function in their regular environment, which is not neuroscience but daily practice. Moreover, it may well be that the same CCC applies to various behaviours that have different causal-neural origins. This is not to deny the reality of beliefs and desires. People *really* believe and desire things. Just like people can *really* be fatigued. A 'belief' or 'a desire' is just as much a *thing* or *process* as 'a fatigue'.

There is much more that can be said about interpretivism, but for our purposes this brief sketch suffices. The most crucial distinction between mental realism and interpretivism with respect to the mind–brain relation – the distinction that makes

[2] Some have argued, influentially, that this option only works if we entirely reduce mental states to brain states (Kim, 1998, 2005), others deny this (Fodor, 1974, 1997). We shall leave this issue aside.

[3] Dennett does not use the CCC/SCC terminology, but he does make the conceptual distinction. CCCs refer to what he calls 'folk-psychological concepts', SCCs refer to concepts that are part of what he calls 'intentional systems theory', which is a scientifically tidied-up version of folk-psychology (Dennett, 1987, pp. 43–68).

all the difference when handling the problems discussed in the Section 2 – is this. Realists think that CCCs/SCCs refer directly to brain processes; the fact that such ascription is based on behaviour – in daily life overt behaviour, in the case of science reaction times, eye-movements, verbal reports, etc. – is a mere *practical* difficulty. Interpretivists, by contrast, acknowledge that the connection between CCCs/SCCs and brain processes is *mediated* by behaviour. CCCs/SCCs interpret, explain and situate behaviour for practical and scientific purposes; in experimental operationalisations, scientists look for the neural processes behind a sub-class of the behaviour to which some CCC or SCC applies. What qualifies as the neural correlate of the CCC/SCC at issue is entirely dependent on the characterization of the relevant behaviour at issue.

This is very significant, for whereas brain processes are objectively there, there is interpretive wiggle room in the characterization of the relevant kind of behaviour in terms of CCCs or SCCs. For instance, when researchers use the Wisconsin card-sorting task (see Fig. 1), they will obtain behavioural data, and perhaps also brain data. How does the researcher know whether to interpret these data as informative for 'task switching' or instead for 'working memory'? The importance of adopting a more interpretivist outlook stems precisely from this observation: the appropriateness of the characterization of the behaviour involved in the operationalisation of a given CCC/SCC is what *turns* the neural causes of that behaviour into a neural correlate of that CCC/SCC. Whether or not a neural process is a correlate of some given CCC or SCC is not an independent, objective fact. The reaction times, eye-movements, verbal reports, etc. that neuroscientists measure in their experiments are not mere indicators of mental processes such as thinking, paying attention, and intending. Rather, the characterization of specific reaction times and eye-movements as instances of 'attention' or 'disgust' or 'intention–detection' is what turns the neural causes of these reaction times or eye-movements in the neural correlates of 'attention', 'disgust' or 'intention–detection'.

4.4. Realism vs. interpretivism in cognitive neuroscience practice: a case study

So how can the view we are advocating be translated to scientific practice? What is the difference between a realist and an interpretivist attitude when it comes to the setting up, interpreting and reporting the outcomes of experiments? In order to answer this question, let us discuss an illustrative example of a scientific paradigm in which CCCs are turned into SCCs, which are in turn operationalised relatively implicitly. In this case the neural activity associated with the SCCs is not directly measured during the task. Rather, the task is used to dissociate two SCCs for which distinct neural correlates are sought through discussion of existing literature. We will first focus on the transition from SCCs to tasks (Fig. 1, arrow b) and then return to the issue of finding the neural correlates for the relevant SCCs (Fig. 1, arrow c).

The paradigm is an experimental set-up used by Lamme in order to argue for the view that visual awareness and attention are distinct processes (Lamme, 2003). It is a change blindness experiment in which an abstract pattern of horizontal and vertical rectangles, arranged in a circle like clock numerals is shown (see Fig. 2). This first stimulus is followed by a grey screen inter stimulus interval followed by the 'same' abstract pattern stimulus in which one rectangle has changed from vertical to horizontal. There are three conditions. In the first condition (see Fig. 2a) the changed rectangle is indicated by a diagonal line, much like a clock hand, during the second stimulus only. Subjects are asked whether they think this rectangle has changed orientation. Overall they perform relatively poorly on this task. In the second condition (see Fig. 2b), the 'clock hand' is present in the first stimulus, indicating the rectangle of which subjects have to determine whether its orientation will be different in the second stimulus. Here, subjects succeed in establishing that the orientation is indeed changed in the second stimulus. The third condition is the interesting one. Here the 'clock hand' is presented in the grey inter stimulus interval, but at exactly the same location in the visual field as the 'clock hands' in the first and second condition (see Fig. 2c). Interestingly, subjects perform almost as good in third condition as they do in the second condition.

The author concludes from this experiment: "Apparently, after the first display had disappeared, a neural representation of almost the whole scene is still present and attention can select from this representation the relevant item in working memory. After the onset of stimulus 2, this representation has vanished, as cueing at that time does not help. The model thus argues for the existence of a short-lived, vulnerable and not easily reportable form of visual experience, which contrasts with a more stable, reportable form of awareness" (Lamme, 2003, p.14). The consciousness of stimulus 1, which is apparently retained during the grey interval, is compared to iconic memory and with what Ned Block calls 'phenomenal consciousness' (Block, 1995). The consciousness that ensues when attention selection is 'added' is compared to working memory or Block's 'access consciousness'.

Lamme is a realist. He treats notions such as phenomenal consciousness, iconic memory, access consciousness, attention and working memory as objectively existing processes. He is keenly aware of the need to defend the distinction between attentional selection and (phenomenal) consciousness. For he recognizes at least three other ways of relating visual awareness to attentional selection in the literature, all of which view attention as a precondition for consciousness. The experiment discussed above is treated as a major clue for the notion that phenomenal consciousness can exist without attention while attention turns it into access consciousness. It is viewed as indirect but nevertheless compelling evidence for an objectively existing distinction between mental capacities. The existence of this distinction is controversial. Many researchers treat the connection between reportability and consciousness as a tautology. Yet this is exactly what the author argues against. For realists there is a fact of the matter as to who is right on this issue.

Interpretivists have a different view on the nature of the conflict. For them, 'attentional selection' is first and foremost a label for an implicit operationalisation: the capacity to single out an item from the first stimulus (or its 'mental after image') and compare it with an item at that location in the second stimulus. The same goes for what is called 'phenomenal

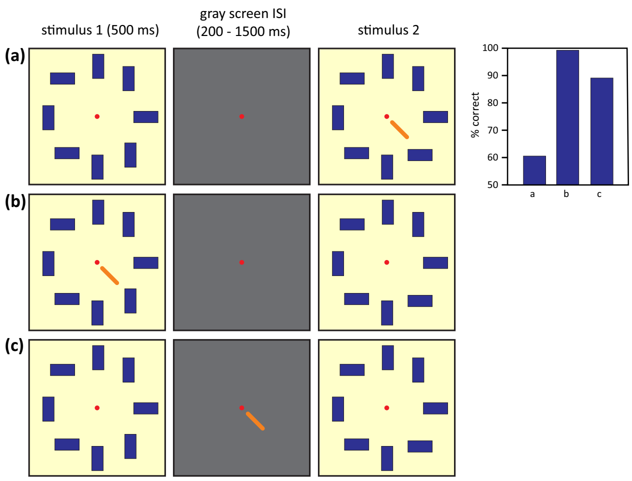

Fig. 2. The experimental paradigm used in the study of Lamme (Lamme, 2003). In the change blindness trials (a–c), a scene containing multiple items is presented (stimulus 1), followed by a grey screen inter-stimulus interval (ISI), after which the same scene (stimulus 2) is shown again. The subject is then asked whether the cued item (indicated by the diagonal line) has changed or not. In (a) it has changed orientation. Subjects perform poorly at this task, (60% correct, lower left histogram). When the to be changed item is cued in advance (b), subjects perform almost 100% correct. However, when subjects are cued after the disappearance of stimulus 1 but before the onset of stimulus 2 (c), they perform almost as well. Reproduced with permission from Lamme (2003).

consciousness' here. That notion is operationalised implicitly as the ability to retain information about the locations and orientations of the rectangles of the first stimulus after the stimulus has gone, for up to 1.5 s. An interpretivist recognizes that what is measured in the experiment are these operationalisations. The terms 'consciousness' and 'attention' are *interpretations* of these operationalisations. This puts the experiment and its contribution to the debate on the relation between attention and visual awareness in a fundamentally different perspective. There are two differences to be highlighted.

First, for an interpretivist the persuasiveness of the experiment as an argument for the separation of phenomenal consciousness and attention depends not just on its outcomes, but just as well on the reasonableness of the interpretations of the implicit operationalisations in terms of the SCCs 'attentional selection' and 'phenomenal consciousness'. Rather than unambiguous evidence for a certain objective state of affairs, the experiment should be seen as persuasive evidence for a specific use of certain SCCs. *Given* that it is reasonable to think of the ability to single out one item from retained information and track its change in the second stimulus as 'attention' and *given* that it is reasonable to think of the retention of the information of the first stimulus as phenomenal consciousness, we should think that phenomenal consciousness does not include attention, based on this experiment.

Thus, interpretivists would demand that for this experiment to be an argument for the separation of consciousness and attention, we need a more elaborate discussion of the reasonableness of the SCCs 'phenomenal consciousness' and 'attentional selection' as interpretations of the operationalisations. Most researchers would agree that 'attentional selection' is a good label for the ability to single out one aspect of the first stimulus or what is retained of that during the interval long enough for it to be compared to its counterpart in the second stimulus. But it seems much more controversial to label the ability to retain information about the location and orientation of the rectangles of the first stimulus during the interval as 'phenomenal consciousness'. The author first interprets the ability to retain this information as a form of iconic memory. The term 'iconic' suggest a kind of internal quasi-visual awareness ("seeing the remembered stimulus with one's minds eye"). And from there the step to phenomenal consciousness is easily made. The point is that strictly speaking we are talking about the retention of information characterized as a form of consciousness through mere associative reasoning. The difference that adopting an interpretivist standpoint would provide is that it would urge for a better argumentation for the label 'phenomenal consciousness', which is crucial for the conclusion in the paper.

A second important difference between the realist and the interpretivist stance is connected with the search for neural correlates. The author takes the above experiment to demonstrate that phenomenal awareness and attentional selection are separate processes. He then undertakes the search for neural correlates for each of these processes by invoking existing neuroscientific knowledge gained in other experimental contexts. Interpretivists stress that experiments do not measure SCCs such as 'attention' or 'consciousness' but the ways in which these SCCs are operationalised in tasks. This means that the neural correlate of some SCC is *in fact* the neural correlate of the capacity to perform that task. What is crucial from an interpretivist standpoint, then, is that the operationalisation of SCCs in the experiment are similar or identical to the operationalisations used in the experiments cited to determine their neural correlates. Combining various studies to draw conclusions about a specific SCC requires a discussion on the similarity or identity of the relevant operationalisations.

Thus, to summarise: an interpretivist's take on the experiments and conclusions presented in this example differs from a realist take in two respects. First, an interpretivist would add a discussion of the characterisations of operationalisation in terms of the SCCs involved. Secondly, an interpretivist would take care that in finding neural correlates of SCCs the same operationalisations are used in psychophysical experiments delineating an SCC from other SCCs and the experiments invoked to determine their neural correlates.

4.5. How cognitive neuroscience could benefit from an interpretivist perspective

Explicit adherence to interpretivism instead of implicit adherence to realism helps to avoid conceptual confusion when it comes to the use of CCCs and SCCs in cognitive neuroscience. Adopting an interpretivist point of view involves acknowledging the interpretive wiggle room that comes with CCCs and to a lesser degree also with SCCs. This implies an awareness of the possibility of conceptual confusion such as the ones discussed in Section 2 and in the case study in Section 4.4. There is nothing wrong with using CCCs and SCCs, as long as it is clear that the operationalisations – i.e., specific behaviour in specific cognitive tasks – are ultimately what is being investigated.

It is certainly possible that a researcher who is implicitly committed to a realist position is well aware of the interpretive translations that are required to investigate cognitive concepts. However, by shifting from a realist to an interpretivist position, one *explicitly* acknowledges that the interpretation of neural processes in terms of SCCs and CCCs is mediated by behaviour. Furthermore, an interpretative stance reminds us that it may well turn out that the CCCs and SCCs we currently employ do not necessarily correspond to any circumscribed neural counterpart (which is emphasized by Buszáki in his quote at the beginning of the article). Finally, this realisation might result in more serious attempts to build cognitive ontologies to formalize scientific cognitive concepts and help to make pragmatic choices between the three strategies discussed in Section 3, should this be required.

We believe the ultimate project of finding the neural mechanisms underlying cognitive functions could benefit significantly from adopting an interpretivist perspective. This would result in knowledge that is easier to integrate and communicate, which is essential for an interdisciplinary field like cognitive neuroscience and for insights that might affect the commonsense understanding of others and ourselves.

5. Conclusion

Popular statements such as that 'your unconscious may be overruling your conscious choices' are fascinating and result in a heated public debate between neuroscientists, philosophers and the wider public. However, such claims are highly problematic in view of the scheme we propose: CCCs such as 'consciousness' are too complex to be broken down in a tractable number of SCCs and concurring operationalisations. Moreover, the tasks involved in most operationalisations only involve parts of SCCs. Despite that, however, popularized neuroscience presents results of experiments based on these partial operationalisations in terms of the original CCC (e.g. 'consciousness'). This, we argue, is unnecessarily misleading. The root of this problem is an implicit adherence to mental realism, the idea that CCCs can be found in the brain. The solution, we argue, is to adopt a more interpretive stance that implies caution when it comes to big conclusions about CCCs. Findings on *aspects* of (un)consciousness and decision making so far are spectacular enough as they are – even without the grander conclusions.

But our aim here is not merely to warn against overly grand conclusions in cognitive neuroscience. It is at least equally important to see that adopting a less realistic and more interpretivist stance toward CCCs helps to become aware of the interpretive nature of the transitions from CCCs to SCCs, from SCCs to cognitive tasks, and from these tasks to the specific neural activity involved. Such awareness may help to avoid significant and detrimental conceptual confusion in an otherwise flourishing branch of science.

Acknowledgments

Thanks to Eelke Spaak, Peter Smittenaar and Tim van Timmeren for helpful comments on a draft of this paper and to Leon de Bruin, Derek Strijbos, Naomi Kloosterboer and Iris van Rooij for fruitful discussion. Thanks to Bas Mooij for help with figure design.

References

Abi-Rached, J. M., & Rose, N. (2010). The birth of the neuromolecular gaze. *History of the Human Sciences, 23*(1), 11–36.
Anderson, M. L., & Pessoa, L. (2011). Quantifying the diversity of neural activations in individual brain regions. In L. Carlson, C. Hölscher, T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 2421–2426). Austin, Texas: Cognitive Science Society.
Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nature Genetics, 25*(1), 25–29.
Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556–559.
Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology, 39*, 15–22.
Bilder, R. M., Sabb, F. W., Parker, D. S., Kalar, D., Chu, W. W., Fox, J., et al (2009). Cognitive ontologies for neuropsychiatric phenomics research. *Cognitive Neuropsychiatry, 14*(4–5), 419–450.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–247.
Buzsáki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.
Campbell, N. (2003). *Mental causation: A nonreductive approach*. New York: Peter Lang.
Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.

Cromwell, H. C., & Panksepp, J. (2011). Rethinking the cognitive revolution from a neural perspective: How overuse/misuse of the term "cognition" and the neglect of affective controls in behavioral neuroscience could be delaying progress in understanding the BrainMind. *Neuroscience and Biobehavioral Reviews, 35*(9), 2026–2035.
Danziger, K. (1997). *Naming the mind: How psychology found its language*. London: SAGE Publications.
Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy, 60*(23), 685–700.
Davidson, D. (1980). *Essays on actions and events*. Oxford: Oxford University Press.
Dennett, D. C. (1971). Intentional systems. *Journal of Philosophy, 68*(February), 87–106.
Dennett, D. C. (1978). *Brainstorms. Philosophical Essays on Mind and Psychology*. Bradford Books.
Dennett, D. C. (1987). *The intentional stance*. Cambridge: MIT Press.
Dennett, D. C. (1991). *Consciousness explained*. Little, Brown.
Fisher, S. E. (2006). Tangled webs: Tracing the connections between genes and cognition. *Cognition, 101*(2), 270–297.
Fodor, J. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese, 28*(2), 97–115.
Fodor, J. (1997). Special sciences: Still autonomous after all these years. In J. Tomberlin (Ed.), *Philosophical perspectives 11, mind, causation, and world* (pp. 149–163).
Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron, 14*(3), 477–485.
Hacking, I. (1986). Making up people. In Heller, Sosna, & Wellberry (Eds.), *Reconstructing individualism* (pp. 222–236). Stanford, California: Stanford University Press.
Han, S., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience, 9*(8), 646–654.
Heil, J., & Mele, A. (1993). *Mental causation*. Oxford: Oxford University Press.
Keefe, R. S. E. (1995). The contribution of neuropsychology to psychiatry. *American Journal of Psychiatry, 152*(1), 6–15.
Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.
Kim, J. (2005). *Physicalism or something near enough*. Princeton: Princeton University Press.
Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9), 2197–2206.
Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7*(1), 12–18.
Lenartowicz, A., Kalar, D. J., Congdon, E., & Poldrack, R. A. (2010). Towards an ontology of cognitive control. *Topics in Cognitive Science, 2*(4), 678–692.
Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123*(1), 3–32.
Macdonald, P. S. (2003). *The history of the concept of mind, volume 1, speculations about soul, mind and spirit from homer to Hume*. Ashgate: Aldershot.
Macdonald, P. S. (2007). *The history of the concept of mind, volume 2, the heterodox and occult tradition*. Ashgate: Aldershot.
Mölder, B. (2009). *Mind ascribed*. Amsterdam: John Benjamins.
Olton, D. S., Becker, J. T., & Handelmann, G. E. (2011). Hippocampus, space, and memory. *Behavioral and Brain Sciences, 2*(03), 313–322.
Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59–63.
Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental States across individuals. *Psychological Science, 20*(11), 1364–1372.
Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics, 5*, 17.
Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology, 22*(3), 262–275.
Putnam, H. (1988). *Representation and reality*. Cambridge: MIT Press.
Sabb, F. W., Bearden, C. E., Glahn, D. C., Parker, D. S., Freimer, N., & Bilder, R. M. (2008). A collaborative knowledge base for cognitive phenomics. *Molecular Psychiatry, 13*(4), 350–360.
Uithol, S., Burnston, D. C., & Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia, 56*, 129–139.
Uttal, W. R. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge: MIT Press.
Van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha- and beta-band oscillations. *The Journal of Neuroscience, 31*(6), 2016–2024.
Vinden, P. G. (2002). Understanding minds and evidence for belief: A study of Mofu children in Cameroon. *International Journal of Behavioral Development, 26*(5), 445–452.