The Mental Lives of Zombies

1. Consciousness and Cognition

A *zombie* is a creature that is just like a conscious subject in all relevant physical,

functional or behavioral respects, except that it has no conscious states – there is

nothing it is like to be a zombie.[1] Zombies have figured prominently in metaphysical

debates about the nature of consciousness, but they can also be usefully employed in

raising questions about the relationship between consciousness and cognition.

Could there be a *cognitive zombie* – that is, a creature with the capacity for cognition

but with no capacity for consciousness? By definition, zombies cannot have

conscious states, but can they nevertheless have cognitive states, such as beliefs, and

cognitive processes, such as reasoning and other forms of rational belief revision?

      In this paper, I am primarily concerned with conceptual questions about the

relationship between consciousness and cognition. As far as possible, I want to

remain neutral on empirical questions about the functional role of consciousness

and metaphysical questions about the nature of consciousness and its place in the

physical world. So, when I ask whether there could be cognitive zombies, the

relevant modality is conceptual possibility, rather than physical possibility or

metaphysical possibility. The question is whether cognitive zombies are

conceptually possible or impossible – that is, whether they can be coherently

conceived or whether this involves some kind of inherent conceptual confusion.[2]

      Without attempting to provide an exhaustive taxonomy, I want to organize

my discussion around three contrasting perspectives on the relationship between

consciousness and cognition. Proponents of *bifurcationism* deny that there is any interesting conceptual connection between consciousness and cognition.[3] In contrast, proponents of *unificationism* agree that there is an interesting conceptual connection between consciousness and cognition, although they disagree about its nature. Some proponents of unificationism endorse a *cognitive analysis of consciousness* on which consciousness is analyzed in terms of its role in cognition.[4] Others endorse a *phenomenal analysis of cognition* on which cognition is analyzed in terms of its relations to consciousness.[5] Thus, proponents of unificationism agree that there is a conceptual connection between consciousness and cognition, although they disagree about whether consciousness is to be analyzed in terms of cognition, or vice versa.

Bifurcationism, which is perhaps the dominant view of the connection between consciousness and cognition, often takes the following form. The concept of consciousness is a *phenomenal* concept that defines consciousness in terms of "what it is like" for the subject. The concept of cognition, on the other hand, is a *functional* concept that defines cognition by its causal role. And crucially, according to proponents of bifurcationism, the phenomenal concept of consciousness is distinct from any functional concept that defines cognition by its causal role.[6]

One symptom of bifurcationism is the widespread view that the problem of explaining consciousness is a *hard problem*, whereas the problem of explaining cognition is an *easy problem*.[7] Explaining consciousness is viewed as a hard problem on the grounds that there is an *explanatory gap* between physical and functional facts on the one hand and phenomenal facts on the other: in other words, it is

2

conceivable that the same physical or functional facts could give rise to different conscious states or none at all. Explaining cognition, by contrast, is viewed as an easy problem on the grounds that there is no such explanatory gap for cognition: it is inconceivable that the same physical or functional facts could give rise to different cognitive states or none at all. Indeed, it is widely held that cognition, unlike consciousness, can be explained in broadly functionalist terms by appealing to the causal properties of physical states. Therefore, proponents of bifurcationism tend to claim that cognition can be explained in functional terms, while denying that consciousness can be so explained.[8]

Unificationism, however, implies that the problem of explaining cognition is intertwined with the problem of explaining consciousness. In that case, the prospects for explaining cognition in physical or functional terms stand or fall with the prospects for explaining consciousness in physical or functional terms. These prospects may seem better or worse depending on the nature of the conceptual connections that are claimed to hold between consciousness and cognition. Given a cognitive analysis of consciousness, we can derive an explanation of consciousness from a prior explanation of cognition, and so the difficulty of explaining consciousness can be ameliorated by the ease of explaining cognition. Given a phenomenal analysis of cognition, on the other hand, an explanation of cognition depends upon a prior explanation of consciousness, and so the problem of explaining cognition inherits the difficulty of explaining consciousness.

Unificationism and bifurcationism have different implications for the question of whether cognitive zombies are coherently conceivable. Bifurcationism

seems committed to the conceivability of cognitive zombies as a consequence of the following two-step argument. First, it is conceivable that there could be a *functional zombie* – that is, a zombie that is a functional duplicate of some conscious creature that has a capacity for cognition. And second, it is inconceivable that a functional zombie, so defined, should lack the capacity for cognition. Therefore, we have the following simple argument for the conceivability of cognitive zombies:

(1) Functional zombies are conceivable.

(2) Functional zombies that lack cognition are inconceivable.

(3) Therefore, cognitive zombies are conceivable.

To illustrate, consider Block's (2002) example of Commander Data, who is functionally just like us, at least at the level of commonsense psychology, but who is physically unlike us insofar as his brain is silicon-based, rather than carbon-based. It seems conceivable that Data is conscious, but it also seems conceivable that he is a zombie; indeed, as Block puts the point, it is an open question whether or not Data is conscious and, moreover, it is a question that we have no conception of how to close. Nevertheless, Block would claim that Commander Data has cognition, since he has states that play the right kind of causal role in the production of behavior.

Unificationism, on the other hand, seems committed to denying that cognitive zombies are coherently conceivable, although the rationale for this will depend upon the nature of the conceptual connections that are claimed to hold between consciousness and cognition. On a cognitive analysis of consciousness,

4

cognitive zombies are inconceivable because functional zombies are inconceivable. Any functional duplicate of a conscious creature with the capacity for cognition is thereby a conscious creature with the capacity for cognition. Therefore, proponents of a cognitive analysis of consciousness accept the second premise of the zombie argument, but they reject the first premise. On a phenomenal analysis of cognition, cognitive zombies are also inconceivable, but for different reasons. Functional zombies are conceivable, but since cognition is analyzed in terms of consciousness, it is inconceivable that any zombie (functional or otherwise) could have cognition. Thus, proponents of a phenomenal analysis of cognition accept the first premise of the zombie argument, but they reject the second premise.

       All three perspectives are represented in the table below:

	Functional zombies?	Cognitive zombies?
Cognitive Analysis of Consciousness	No	No
Phenomenal Analysis of Cognition	Yes	No
Bifurcationism	Yes	Yes

With these options in mind, let me outline the goals of the paper. The main goal is to argue for the incoherence of cognitive zombies by appealing to a phenomenal analysis of cognition, as opposed to a cognitive analysis of consciousness. Indeed, for present purposes, I will set aside the cognitive analysis of consciousness and work on the assumption (in my view, a plausible one) that functional zombies are coherently conceivable. The task remains to argue against bifurcationism by

5

articulating, defending and motivating a phenomenal analysis of cognition on which it is inconceivable that functional zombies have cognition.

Although I am not the first to propose a phenomenal analysis of cognition, my arguments are different from the familiar ones. The usual strategy is to argue that the connection between consciousness and cognition is a consequence of a more fundamental connection between consciousness and intentionality. As I will explain, however, there are good empirical reasons for rejecting the proposed connection between consciousness and intentionality. My strategy, in contrast, is to argue that the connection between consciousness and cognition can be derived from a more fundamental connection between consciousness and rationality. As such, this paper forms part of a much larger project of exploring the normative role of consciousness and its implications for our mental lives.[9]

2. Two Conceptions of Cognition

Could there be a cognitive zombie? Let us suppose, at least for the sake of the argument, that there could be a functional zombie – that is, an unconscious creature that is a functional duplicate of a conscious creature with a capacity for cognition. It is a further question whether functional zombies have cognition. Do functional zombies have cognitive states, such as beliefs, and cognitive processes, such as reasoning and rational belief revision? Perhaps not surprisingly, the answer to this question depends upon how we understand the concept of cognition. In this section, I draw a distinction between two conceptions of cognition – the *behavioral* analysis

and the *phenomenal* analysis – that have different implications for the question of whether functional zombies have cognition.

2.1. The Behavioral Analysis

On the *behavioral analysis of cognition*, beliefs and other cognitive states are analyzed in terms of their dispositions to cause physical behavior. The behavioral analysis comes in simpler and more sophisticated versions. In particular, we can distinguish between *analytical behaviorism*, which is the earliest and simplest incarnation of the behavioral analysis of cognition, and *analytical functionalism*, which is a more sophisticated descendent of analytical behaviorism.[10]

According to analytical behaviorism, all psychological concepts, including the concept of belief, can be given a reductive analysis in terms of behavioral dispositions described in wholly non-psychological terms. On this view, there are conceptual truths of the form: one has a belief B if and only if one has a disposition D to engage in physical behavior P in circumstances C. The standard objection to this view is that the analysis is circular, since we cannot specify the circumstances in which beliefs manifest themselves in behavior without making reference to the presence or absence of background beliefs. Given the holistic relationship between beliefs and behavioral dispositions, we cannot analyze beliefs in terms of their behavioral dispositions alone.

Analytical functionalism retains the same ambition to give a reductive analysis of psychological concepts such as belief in terms of non-psychological concepts such as physical behavior. However, it avoids the circularity problem by

giving an analysis of beliefs and other psychological states in terms of their place in

a causal network that grounds behavioral dispositions holistically, rather than

atomistically. As Lewis (1972) explains, we begin with an explicit statement of the

causal relationships between psychological states, environmental inputs and

behavioral outputs using the terms of our commonsense psychological theory. Next,

we generate the *Ramsey sentence* for the theory by eliminating all the psychological

terms in the theory and systematically replacing them with existentially quantified

bound variables. The result is a complex definite description that specifies the

causal role of beliefs and other psychological states in non-psychological terms.

Block (1978) and others have argued that analytical functionalism fails as an

analysis of consciousness. There seems to be no incoherence in the idea that a

zombie that satisfies the Ramsey sentence for commonsense psychology. All we

need is a functional isomorphism that maps states of a conscious creature onto

states of the zombie in a way that preserves causal relations between them.

Nevertheless, Block does not reject functionalism altogether. Instead, he

recommends the *containment response* – that is, to abandon the functionalist

analysis of consciousness, while retaining the functionalist analysis of cognition.[11]

On this proposal, there could be a cognitive zombie, since satisfying the Ramsey

sentence for commonsense psychology is sufficient for cognition, although it is not

sufficient for consciousness.[12]

Despite Block's recommendation, however, it is not clear that we should

accept this containment response. Why should we suppose that the functional role

of cognition can be specified without appealing to phenomenal consciousness? After

8

all, beliefs are dispositional states that manifest themselves not only in physical

behavior, but also in phenomenal consciousness. So why privilege behavioral

dispositions, as opposed to phenomenal dispositions, in the analysis of cognition?

In historical terms, the behavioral analysis of psychological concepts seems

to have been motivated primarily by verificationist assumptions about semantics

together with a desire to avoid certain metaphysical and epistemological problems

about consciousness, such as dualism and skepticism about other minds.[13] However,

the containment response cannot be motivated in this way, since it eschews any

behavioral or functional analysis of consciousness. If the metaphysical and

epistemological problems of consciousness can be solved, then there is no need for

containment, but if they cannot be solved, then containment will not be enough to

make them go away. So, unless the containment response can be motivated on other

grounds, the suspicion will remain that it is an ideological relic of a previous era.

2.2. The Phenomenal Analysis

On the *phenomenal analysis of cognition*, beliefs and other cognitive states are

analyzed in terms of their dispositions to cause phenomenally conscious episodes of

judgment, rather than their dispositions to cause physical behavior. Unlike the

behavioral analysis, the phenomenal analysis makes no attempt to give a reduction

of psychological concepts in non-psychological terms. This removes an obstacle that

otherwise prevents one from appealing to phenomenal consciousness in an analysis

of cognition. In this section, I will articulate a phenomenal analysis of cognition and

defend it against some preliminary objections, but I will defer the task of motivating this analysis until later sections of the paper.

I begin with a distinction between belief and judgment. As I use the term, judgments are phenomenally conscious episodes that have phenomenal character: there is something it is like for a subject to judge that a proposition is true. Moreover, judgments are individuated by their phenomenal character: two judgments are judgments of the same kind if and only if they have the same phenomenal character. Beliefs, unlike judgments, have no phenomenal character: there need be nothing it is like for a subject to believe that a proposition is true. Nevertheless, beliefs are disposed to cause phenomenally conscious judgments; indeed, they are individuated by these dispositions: two beliefs are beliefs of the same kind if and only if they have are disposed to cause judgments with the same phenomenal character.[14]

Just as there is a distinction between outright belief and outright judgment, so there is a distinction between degrees of belief and their manifestations in phenomenal consciousness as feelings of confidence in a given proposition. One's feelings of confidence are phenomenally conscious episodes that are individuated by their phenomenal character, whereas one's degrees of belief are individuated by their dispositions to cause the phenomenal character of feelings of confidence. In this way, the phenomenal analysis of outright belief can be plausibly extended to degrees of belief too.

Beliefs and judgments are *intentional* states – they are intentional attitudes towards intentional contents that specify the conditions under which they are true.

Moreover, it is widely assumed that beliefs and judgments are individuated by their intentional properties – that is, by their intentional contents and their intentional attitude-types. However, we need not choose between individuating beliefs and judgments in terms of their intentional properties, as opposed to their phenomenal properties, or the phenomenal properties that they are disposed to cause. Given plausible assumptions, these are equivalent ways of individuating them.

The key assumption is that the phenomenal character of judgment is sufficient to determine its intentional content and its intentional attitude-type. On this assumption, the phenomenal character of judgment is *content-specific* and *attitude-specific* in the sense that what it is like to judge a proposition is different from what it is like to adopt the same attitude towards a different proposition and what it is like to judge a proposition is different from what it is like to adopt a different attitude towards the same proposition. This assumption is not uncontroversial, but it has been widely discussed elsewhere, and so for present purposes, I will simply take it for granted.[15]

One important objection to the assumption above is that the intentional contents of beliefs and judgments are externally individuated by their relations to the environment. On this view, there can be phenomenal duplicates whose beliefs and judgments have different intentional contents in virtue of their different relations to the environment. However, this is consistent with the claim that the beliefs and judgments of phenomenal duplicates also share intentional contents in virtue of their shared phenomenal character. Thus, we can distinguish between the *narrow* contents of belief and judgment that are individuated by their phenomenal

character alone and the *wide* contents of belief and judgment that are individuated by their phenomenal character together with their relations to the environment.[16]

Another objection to the phenomenal analysis of cognition is that there are familiar cases of self-deception in which one's dispositions towards judgment come apart from one's dispositions to engage in physical behavior. Moreover, in such cases, there is some tendency to suppose that our beliefs are reflected in our actions, rather than our judgments. For instance, Peacocke (1998) gives the example of an academic on a hiring committee who is disposed to judge that foreign degrees are equal in standard to domestic degrees, although her votes in hiring decisions reveal that she does not really believe this; indeed, what she really believes is that foreign degrees are inferior, although this is not what she is disposed to judge. So described, the case presents a counterexample to the phenomenal analysis of belief.

My response is to dispute this description of the case. One possibility is that the academic has inconsistent beliefs. In some contexts, such as hiring decisions, she may be disposed to judge that foreign degrees are equal, while in other contexts, such as the local tavern, she may be disposed to judge that foreign degrees are inferior. We can explain away any lingering temptation to say that what she *really* believes is that foreign degrees are inferior, since this is the belief that is primarily operative in guiding her behavior in hiring decisions.

Alternatively, perhaps there are no suitable contexts in which the academic is disposed to judge that foreign degrees are inferior to domestic degrees. In that case, she does not believe that foreign degrees are inferior, although she behaves in many respects *as if* she does.[17] In that case, her rational defect is practical, rather than

epistemic. We may assume that her beliefs and judgments are rationally formed in a way that is appropriately responsive to the evidence. Her problem is not that her beliefs are unjustified, but rather that her behavior is unjustified, since her justified beliefs fail to exert an appropriate influence on her behavior.

We can imagine cases in which one's behavioral dispositions are even more radically dissociated from one's dispositions towards judgment. For instance, Strawson (1994) gives the following example of the Weather Watchers:

> The Weather Watchers are a race of sentient, intelligent creatures. They are distributed about the surface of their planet, rooted to the ground, profoundly interested in the local weather. They have sensations, thoughts, emotions, beliefs, desires. They possess a conception of an objective, spatial world. But they are constitutionally incapable of any sort of behavior, as this is ordinarily understood. They lack the necessary physiology. Their mental lives have no other-observable effects. They are not even disposed to behave in any way. (1994: 251)

Strawson claims that the Weather Watchers have beliefs about the local weather, although their beliefs do not dispose them to engage in physical behavior at all. According to the behavioral analysis, of course, the case is incoherent as described, since beliefs are analyzed in terms of their behavioral dispositions. According to the phenomenal analysis, however, the case is perfectly coherent, since beliefs are

analyzed in terms of their phenomenal dispositions, which are merely contingently associated with their behavioral dispositions.

These two conceptions of cognition – the behavioral analysis and the phenomenal analysis – also have different implications for the conceivability of cognitive zombies. If cognition is analyzed in terms of its phenomenal dispositions, as the phenomenal analysis claims, then cognitive zombies are incoherent. But if cognition is analyzed in terms of its behavioral dispositions, as the behavioral analysis claims, then cognitive zombies are perfectly coherent. What basis do we have, then, for deciding between these competing conceptions of cognition?

2.3. A Terminological Debate?

Our ordinary concept of belief is associated with a cluster of dispositions, including dispositions to make judgments, dispositions to engage in behavior, and perhaps other dispositions besides. These dispositions usually come together, but there are conceivable scenarios, such as zombie scenarios, in which they come apart. Do we have any principled reasons for privileging some of the dispositions in the cluster, rather than others, for purposes of analyzing the concept of belief? Or is this merely to insist on a terminological stipulation about how to use the word 'belief'?

Schwitzgebel (2002, 2010) avoids the choice between the phenomenal analysis and the behavioral analysis by opting instead for a cluster analysis on which one has a belief just in case one has enough of the dispositions in the cluster that is associated with our ordinary use of the term 'belief'. On this view, there are cases of "in-between believing" in which it is indeterminate whether or not one has

a belief because one has some but not all of the dispositions in the cluster. Presumably, for instance, it is indeterminate whether or not zombies have beliefs, since they have the behavioral dispositions, but not the phenomenal dispositions, that are associated with the concept of belief. Of course, we might sharpen the concept of belief in such a way as to deliver a determinate answer to the question of whether or not zombies have beliefs. However, the question remains whether there is any substantive question, as opposed to a merely terminological one, about whether we should precisify the concept of belief in one way rather than another.

Chalmers (1996) draws a distinction between two concepts of mind – the phenomenal concept of mind as conscious experience and the psychological concept of mind as the causal-explanatory basis of behavior. Moreover, Chalmers claims that there is no substantial question, as opposed to a merely terminological question, about which of these is the correct analysis of mind:

> On the phenomenal concept of mind, mind is characterized by the way it *feels*; on the psychological concept, mind is characterized by what it *does*. There should be no question of competition between these two notions of mind. Neither of them is *the* correct analysis of mind. They cover different phenomena, both of which are quite real. (1996: 11)

The distinction between phenomenal and behavioral (or psychological) conceptions of cognition is a special case of Chalmers' distinction between two concepts of mind. Accordingly, perhaps there is no substantial question, as opposed to a merely

terminological question, about which is the correct analysis of cognition. Everyone

can agree that zombies have beliefs in the behavioral sense, but not in the

phenomenal sense. The only question that remains is whether we should use the

term 'belief' to express the phenomenal concept or the behavioral concept of belief,

but this seems like a merely terminological question, rather than a substantive one.

In more recent work, Chalmers (2011) makes a useful methodological

proposal about how to avoid terminological debates of this kind. The proposal is

that we should begin by clarifying the theoretical roles that we associate with the

concept of belief, and then ask what belief must be like in order to play that role:

> On the picture I favor, instead of asking, 'What is X', one should focus on the
>
> roles one wants X to play, and see what can play that role. ... [I]nstead of
>
> asking, "What is a belief? What is it to believe?" and expecting a determinate
>
> answer, one can instead focus on the various roles one wants belief to play,
>
> and say, here are some interesting states: B_1 can play these roles, B_2 can play
>
> these roles, B_3 can play these roles. Not much hangs on the residual
>
> terminological question of which is really belief. (2011: 538)

I agree wholeheartedly with this proposal. We need not concern ourselves with

questions about our ordinary use of the term 'belief', such as whether we use it in a

way that expresses the phenomenal concept, the behavioral concept, or some

cluster concept. Instead, we should ask which of these concepts are most useful for

us to have because they pick out states that play an important theoretical role.

Moreover, we should be open to a form of pluralism on which our ordinary concept of belief conflates a range of useful concepts, each of which picks out a unique kind of state that plays an important and distinctive theoretical role.

My goal in this paper is to argue that the phenomenal concept of belief picks out a kind of state that plays an important and distinctive theoretical role. I do not intend to deny that there are other concepts in the vicinity that pick out more generic or more specific kinds of states that also play an important and distinctive theoretical role. My claim is simply that there is an important theoretical distinction in the vicinity of our ordinary concept of belief that cannot be drawn except in terms of phenomenal consciousness. This is a substantive claim, and not merely a terminological one, which many philosophers are inclined to reject. Indeed, Chalmers himself claims that we can "subtract out" the phenomenal component of belief without thereby sacrificing anything of theoretical importance:

> The most substantial requirements for having a specific belief will lie elsewhere than in the phenomenal. One could even subtract any phenomenal component out, leaving a concept of pseudobelief that resembles belief in most important respects except that it does not involve the concept of consciousness. Indeed, it is plausible that pseudobelief could do most of the explanatory work that is done by the concept of belief. (1996: 20)

Chalmers' claim here is that the phenomenal concept of belief does not correspond to any theoretically interesting kind of state and so, for theoretical purposes, it is the

behavioral (or psychological) concept, rather than the phenomenal concept, that we should be interested in. The challenge for proponents of the phenomenal analysis is to show that there is some theoretically important role that a belief can play if and only if it is individuated by its relations to phenomenal consciousness.

In what follows, I consider two ways of arguing for the theoretical significance of phenomenal consciousness in an analysis of cognition. The first strategy is to argue that the connection between consciousness and cognition is a consequence of a more fundamental connection between consciousness and intentionality. However, I argue that there are good empirical reasons for rejecting the proposed connection between consciousness and intentionality. Instead, I propose an alternative strategy on which the connection between consciousness and cognition is explained as a consequence of a more fundamental connection between consciousness and rationality.

3. Consciousness and Intentionality

What is the relationship between consciousness and intentionality? Many contemporary philosophers endorse a weak version of *intentionalism*, according to which all conscious states are intentional states:

Intentionalism: all conscious states are intentional states.

Some philosophers also endorse the converse of intentionalism, which we might call the *intentional connection thesis*.[18] In its strongest version, the intentional

connection thesis says that all intentional states are conscious states. But there is also a weaker version, which says that all intentional states are either conscious states or they are individuated by their relations to conscious states:

The intentional connection thesis: all intentional states are either conscious states or they are individuated by their relations to conscious states.

If the intentional connection thesis is true, then all intentional states are states of conscious creatures and so there cannot be an *intentional zombie*.[19] Moreover, assuming that all cognitive states are intentional states, it follows that all cognitive states are states of conscious creatures and so there cannot be a *cognitive zombie*. In this way, the connection between consciousness and cognition can be derived from a more fundamental connection between consciousness and intentionality:

(1) All cognitive states are intentional states.

(2) All intentional states are either conscious states or individuated by their relations to conscious states.

(3) So, all cognitive states are either conscious states or individuated by their relations to conscious states.

The first premise seems uncontroversial, but the second premise stands in need of further argument. Why should we accept the proposed connection between intentionality and consciousness?

Perhaps the most influential line of argument is that the intentional connection thesis is needed in order to explain the intentionality of cognition in a way that avoids the problem of radical indeterminacy.[20] There are well known attempts to explain the intentionality of cognition in broadly causal terms by appealing to the functional role of intentional states in cognition together with their causal relations to environmental inputs and behavioral outputs. However, these causal theories of intentionality face equally well known problems explaining how the intentionality of cognition is not radically indeterminate.[21]

Some argue that the problem of radical indeterminacy arises because of a failure to appreciate the role that consciousness plays in securing the determinate intentionality of cognition. The suggestion is that conscious states have determinate intentionality in virtue of their determinate phenomenal character, while unconscious states have determinate intentionality in virtue of their relations to the determinate phenomenal character of phenomenally conscious states. This yields the following argument for the intentional connection thesis:

(1) All intentional states have determinate intentional properties in virtue of being conscious or individuated by their relations to conscious states.

(2) So, all intentional states are either conscious states or individuated by their relations to conscious states.

I will argue, however, that the argument rests on a false premise: we have good empirical reasons to suppose that some intentional states have determinate

intentional properties, although they are neither conscious nor individuated by their relations to consciousness.

In rejecting the premise, of course, one takes on the burden of explaining intentionality in a way that solves the problem of radical indeterminacy, while avoiding any appeal to consciousness. This is a heavy burden. After all, it is not an attractive option to take intentionality as a primitive and irreducible feature of the world.[22] And yet nobody has succeeded in showing how intentionality can be reductively explained in purely causal terms in such a way as to avoid the problem of radical indeterminacy. Following Lewis (1983), it may be that we can give a reductive explanation of intentionality that avoids radical indeterminacy if we supplement the appeal to causation with further assumptions about the eligibility of objects in the world to serve as the objects of our intentional states. In any case, I assume that one of these options must be viable, since we have good empirical reasons to reject the intentional connection thesis.

Let us begin with the strongest version of the intentional connection thesis, which states that all intentional states are conscious states. This version of the intentional connection thesis conflicts with the explanatory role that unconscious intentional states play in commonsense psychology and scientific psychology alike. For instance, commonsense explanations of action appeal to unconscious beliefs, desires and intentions of the agent, while computational explanations in cognitive science appeal to computational processes defined over unconscious mental representations, such as Chomsky's (1965) tacit knowledge of syntax and Marr's (1982) primal, 2.5D, and 3D sketch.

Some proponents of the intentional connection thesis claim that, strictly speaking, there is no unconscious intentionality, although it can be useful to speak metaphorically as if there were.[23] However, unconscious intentionality seems to play an indispensable role in psychological explanation in common sense and cognitive science. Moreover, we have good reasons to believe in existence of these states, rather than regarding them as useful fictions, insofar as they play an indispensable role in explanation. After all, it is a widely accepted methodological precept that we ought to believe in the entities posited by our best theories. Therefore, we should endorse realism, rather than eliminativism or instrumentalism, about the existence of unconscious intentional states.

Others weaken the intentional connection thesis in an attempt to make it consistent with the existence of unconscious intentionality. The general idea is that all unconscious intentionality is individuated by its relations to consciousness. Consider Searle's (1990) connection principle:

> The ascription of an unconscious intentional phenomenon to a system implies that the phenomenon is in principle accessible to consciousness. (1990: 333)

According to Searle, an unconscious intentional phenomenon is "accessible to consciousness" just in case it is "potentially conscious" in the sense that it is "a possible conscious thought or experience" (1990: 336). This prompts the objection that beliefs are not accessible to consciousness, since they are distinct from the

potentially conscious manifestations that they cause. However, there is a more plausible variation on Searle's proposal, according to which the *contents* of beliefs are accessible to consciousness as the contents of conscious judgments. This proposal is best understood as a consequence of the claim that beliefs are individuated wholly by their dispositions to cause judgments with a certain content-specific phenomenal character. As Searle acknowledges, however, this proposal cannot be extended to the "subdoxastic" mental representations that figure in computational explanations in cognitive science.[24]

Subdoxastic states, unlike beliefs, are not individuated wholly by their disposition to cause phenomenally conscious states. On the contrary, they are individuated at least in part by their dispositions to play a role in computational processes that occur below the level of phenomenal consciousness. To illustrate the point, consider Davies' (1989) hypothetical example of states of tacit knowledge of language that are disposed to cause phenomenally conscious itches or tickles. Presumably, what makes it the case that these states embody tacit knowledge of language is not their disposition to cause itches and tickles, but rather their roles in linguistic processing.

Consider Quine's (1970) challenge to Chomsky's notion of tacit knowledge. Quine's challenge is to explain what constitutes tacit knowledge of a rule if it is less demanding than explicit knowledge of the rule, but more demanding than merely exhibiting linguistic behavior that conforms to the rule. The standard account is that tacit knowledge of a rule is a matter of the causal structure in the psychological processes that underpin one's linguistic behavior.[25] More specifically, one has tacit

knowledge of a rule if and only if the causal structure of one's psychology mirrors the logical structure of a theory that includes that rule. There could be two subjects that exhibit the same linguistic behavior, although their behavior is explained by psychological processes that embody tacit knowledge of different linguistic rules. Thus, tacit knowledge is individuated not merely by its disposition to cause linguistic behavior, but also by its role in psychological processes that occur beneath the level of phenomenal consciousness.

This point can be generalized to other subdoxastic mental representations, including those involved in vision. There could be two subjects that have the same visual experiences, although their visual experiences are explained by different kinds of visual processing involving different representations and rules. Thus, visual representations and rules are individuated not just by their role in explaining conscious experience, but also by their role in psychological processing that occurs beneath the level of phenomenal consciousness.

One reaction would be to weaken the intentional connection thesis even further so as to allow for subdoxastic mental representations that are individuated in part by their role in unconscious computational processes, but also in part by their relations to conscious states. After all, this weakened version of the intentional connection thesis is strong enough to sustain the claim that all intentional states are states of conscious subjects and hence that there cannot be an intentional zombie. For instance, Horgan and Graham (2012) claim that consciousness is an "anchor point" for intentionality in the sense that all unconscious intentional states are causally integrated within a network that includes conscious intentional states. This

24

seems plausible for some, but not all, unconscious intentional states. Some visual representations are constrained by their role in explaining the intentional properties of conscious visual experience, but there are others that play no role in explaining the intentional properties of conscious visual experience and which are individuated wholly by their role in explaining behavior.

To illustrate the point, consider Milner and Goodale's (1995) distinction between two visual streams: the ventral stream processes visual information for the conscious identification and recognition of objects, while the dorsal stream processes visual information for the unconscious control of action. This hypothesis is designed to explain a range of empirical data, including cases in which one's experience of an object conflicts with the way in which one acts upon the object. In the Titchener illusion, for example, subjects whose experience misrepresents the size of a coin are able to accurately proportion the size of their grip in reaching for the coin. In such a case, one's experience of the coin is explained by representations in the ventral stream that misrepresent the size of the coin, while the spatial parameters of one's visually guided action are explained by representations in the dorsal stream that accurately represent the size of the coin.

If this hypothesis is correct, then visual representations in the ventral stream are individuated in part by their role in explaining the conscious identification and recognition of objects, whereas visual representations in the dorsal stream are individuated wholly by their role in the unconscious control of action. Of course, there are some functional connections between the dorsal stream and the ventral stream, but these connections are highly circumscribed. Moreover, if vision evolved

in response to motor demands, as Milner and Goodale claim, then we should expect to find phylogenetically more ancient creatures that have some analogue of the dorsal stream, but no analogue of the ventral stream. Presumably, for instance, honey bees and desert ants have perceptual systems that explain the complex, relational properties of their behavior, although it is an open question whether or not these creatures are conscious.[26]

I conclude that the intentional connection thesis is probably not true and, in any case, it is certainly not conceptually true. It is conceivable that some intentional states are neither conscious nor individuated by their relations to consciousness. Indeed, it is conceivable that some intentional states are states of unconscious creatures. Zombies, like the zombie systems within us, are best construed as having unconscious intentional states that explain the relational properties of behavior. Therefore, we must conclude that there can be intentional zombies after all.

4. Consciousness and Rationality

If intentionalism is true, then all conscious states are intentional states, but if the intentional connection thesis is false, as I argued in the previous section, then not all intentional states are conscious or individuated by their relations to conscious states. Nevertheless, there may be an important theoretical distinction to be drawn between intentional states depending on their relations to consciousness. Various philosophers have endorsed the idea that consciousness is the basis of an important theoretical distinction between intentional states, including Davies (1995) and Campbell (2002), who summarizes the point like this:

The cautious view...is that we have (at least) two different types of

representation (Davies 1995). On the one hand, there are the conceptual

contents of ordinary beliefs and desires, to which consciousness may

constitutively attach. On the other hand, there are the non-conceptual

contents of information-processing states. (2002: 12)

Meanwhile, others remain highly skeptical about the theoretical significance of any

distinction between intentional states that is drawn in terms of consciousness. Thus,

for instance, Chomsky (1975) writes:

It may be expected that conscious beliefs will form a scattered and probably

uninteresting subpart of the full cognitive structure. (1975: 163)

Meanwhile, Fodor (1975) issues the following challenge for proponents of the

theoretical significance of consciousness:

That will depend upon whether there *are* generalizations which hold (just)

for conscious mental states, and that depends in turn on whether the

conscious states of an organism have more in common with one another than

with the *un*conscious states of the nervous system of the organism. It is, in

this sense, an open question whether conscious psychological states provide

a natural domain for a theory. (1975: 52, n.19)

The challenge is to find some important theoretical generalizations that hold just for intentional states that are either conscious or individuated by their relations to consciousness. Fodor himself remains unconvinced that this can be done.

In this section, I respond to Fodor's challenge by proposing that there is an important theoretical connection between consciousness and rationality. More specifically, I propose to replace the intentional connection thesis with the following rational connection thesis:

> The rational connection thesis: an intentional state plays a rational role if and only if it is either conscious or individuated in such a way that its content is accessible to consciousness as the content of a conscious state.

If the intentional connection thesis is false, then not all intentional states are conscious or individuated by their relations to conscious states. But if the rational connection thesis is true, then there is an important theoretical distinction to be drawn between those intentional states that satisfy this criterion and those that do not. More importantly, for current purposes, the rational connection thesis can be used as a premise in arguing for the phenomenal analysis of cognition:

(1) All cognitive states are intentional states that play a rational role.

(2) All intentional states that play a rational role are either conscious or individuated by their relations to conscious states.

(3) So, all cognitive states are either conscious or individuated by their relations to conscious states.

This argument derives the connection between consciousness and cognition from a more fundamental connection between consciousness and rationality together with rational constraints on cognition. According to this argument, there cannot be a *cognitive zombie* because there cannot be a *rational zombie.*

The rationale for the first premise is that cognition is the domain of beliefs, which unlike subdoxastic mental representations, play a rational role in reasoning. Some, like Chomsky (1975), use the term 'cognition' in a broader sense on which all intentional states and processes count as cognitive states and processes. As I use the term, by contrast, cognitive states and processes are distinguished from merely computational states and processes by virtue of their rational role in reasoning. What is crucial here is not the terminological issue of whether the term 'cognition' should be used in a broad sense or a narrow sense, but rather the substantive claim that there is a theoretically important distinction to be drawn between intentional states and processes that play a rational role and those that do not.

What does it mean to say that all cognitive states play a rational role? We can distinguish between weak and strong versions of this rationality constraint on cognition. The strong version of the rationality constraint says that all cognitive states figure in cognitive processes that are either rational or approximately rational. However, this version is too strong to be plausible: there is substantial empirical evidence to suggest that many of our cognitive processes are neither

29

rational nor even approximately rational.[27] Nevertheless, there is a weak version of the rationality constraint that is perfectly consistent with the empirical evidence. This version says that all cognitive states provide a source of reasons (or perhaps rational requirements) to engage in some cognitive processes, rather than others. The weak version of the rationality constraint does not entail that all cognitive processes are rational, or even approximately rational, but only that they are subject to rational assessment.

Cognitive processes and computational processes are in many ways alike. Many computational processes, like cognitive processes, are intentional processes that involve causal transitions between intentional states that are causally sensitive to their intentional properties.[28] The difference is that computational processes, unlike cognitive processes, are not subject to rational assessment: it makes no sense to ask whether they are reasonable or rational. This point is best illustrated by means of examples.

Chomsky (1965) explains our ability to understand syntactically well-formed sentences of our native language in terms of our tacit knowledge of a syntactic theory for the language. For instance, when we understand the syntactic structure of a sentence, this is explained as the result of an inference-like computational process defined over representational states that mirror the steps in a logical derivation of the syntactic structure of the sentence.

Similarly, Marr (1982) explains visual experience of the environment in terms of inference-like computations defined over a series of representational states, including the primal sketch, the 2.5D sketch, and the 3D sketch. These

30

computations take as input the representation of light intensity in a pair of two-dimensional retinal images and yield as output a three-dimensional representation of shapes and their spatial organization in an object-centered frame of references.

These computational processes are *inference-like* in the sense that they are transitions between intentional states that are sensitive to the intentional properties of those states. Unlike genuine inferences, however, they are not subject to rational assessment. It makes no sense to ask whether it is reasonable or rational for the syntactic system to compute the syntactic structure of a sentence. Likewise, it makes no sense to ask whether it is reasonable or rational for the visual system to compute a representation of the distal environment from a pair of retinal images. These intentional processes are simply not subject to rational assessment at all.

That is not to say that these intentional processes are not subject to any other kinds of normative assessment. The point is rather that the normative standards that are relevant to the assessment of computational processes are distinct from the normative standards that are relevant to the assessment of cognitive processes. For current purposes, the distinction can be left at a more or less intuitive level. Ultimately, of course, it would be desirable to give a more theoretical account of the distinction between computational and rational norms. This is beyond the scope of the present paper, but I have argued elsewhere that rational norms are subject to structural principles of accessibility that need not apply in the computational domain.[29]

Assuming that there is a theoretically important distinction to be drawn between the norms that govern our cognitive processes and the norms that govern

our computational processes, we can ask what explains this distinction. Why are cognitive and computational processes subject to different kinds of normative assessment? Presumably, the normative distinction between these intentional processes cannot simply be taken as primitive; rather, it should be explained in terms of some non-normative distinction between the intentional states that figure in those intentional processes. In particular, the normative distinction between cognition and computation should be explained in terms of some non-normative distinction between cognitive states and computational states – that is, between beliefs and subdoxastic states. But what is the nature of this distinction?

One tempting avenue is to appeal to Dennett's (1969) distinction between personal and subpersonal levels. On this proposal, beliefs are intentional states of the person, whereas subdoxastic states are intentional states of parts of the person – namely, their computational subsystems. The problem with this proposal is that we need a more fundamental account of what makes it the case that an intentional state is properly attributed to the person as opposed to one of the person's subsystems. Bermudez (1995) makes the point effectively:

> Either personal level states have further features in virtue of which they are properly attributable to persons rather than parts of persons, or they do not. If they do not then the distinction is doomed. But if they do have such further features then it makes more sense to state the distinction in terms of these further features. (1995: 353)

Broadly speaking, there are two options for cashing out the distinction between personal-level and subpersonal-level intentional states: one can appeal either to facts about consciousness or to facts about functional role.

The problem with appealing to consciousness is that not all of our cognition occurs within the stream of consciousness. On the contrary, much of our reasoning and rational belief revision draws on background information that is represented unconsciously in the belief system. For instance, inductive reasoning is causally sensitive to vast amounts of background information, not all of which can be brought to consciousness in the process of drawing a conclusion from observed evidence. Indeed, there may be inferential processes that occur entirely below the level of consciousness – as when one realizes that one has discovered the solution to a problem without consciously thinking about it. So how can we explain the normative distinction between unconscious cognitive processes and computational processes by appealing to consciousness alone?

The answer is that we can explain the normative distinction between cognition and computation by appealing to the rational connection thesis in combination with a phenomenal analysis of cognition. According to the rational connection thesis, an intentional state plays a rational role only if it is either conscious or individuated by its relations to conscious states in such a way that its content is accessible to consciousness as the content of a conscious state. Moreover, according to the phenomenal analysis of cognition, cognitive states are either conscious or individuated in the right way by their relations to conscious states, whereas computational states are individuated at least in part by their role in

unconscious computational processes. Together, these two claims explain the datum that cognitive states play a rational role, whereas computational states do not. Thus, we have an argument by inference to the best explanation for a package that explains the rationality of cognition in terms of a phenomenal analysis of cognition together with a connection between rationality and consciousness.

5. Consciousness and Functional Role

In the previous section, I argued that the normative distinction between cognition and computation is explained by a phenomenal analysis of cognition together with a connection between rationality and consciousness. But this argument, like any inference to the best explanation, is vulnerable to the objection that there are alternative, and superior, explanations of the datum to be explained. In this context, the obvious alternative is to explain the normative distinction between cognition and computation by appealing to some functional analysis of cognition together with a connection between rationality and functional role. So, the challenge for proponents of the functional analysis of cognition is to develop a plausible explanation of the normative distinction between cognition and computation in purely functional terms. In this section, I consider what is, in my view, the most promising version of a functional explanation of this kind and I argue that it fails. I conclude, pending further proposals, that the phenomenal analysis of cognition is to be preferred to a functional analysis of cognition on the grounds that it provides a better explanation of the rationality of cognition.[30]

One influential proposal associated with Stich (1978) and Fodor (1983) is that the distinction between computation and cognition can be explained in terms of modularity. According to this proposal, computation is functionally isolated within a series of distinct, modular subsystems, whereas cognition is functionally integrated within a unified, nonmodular system.

For instance, Stich (1978) claims that beliefs are functionally integrated with one another by means of their inferential connections, whereas subdoxastic states are functionally isolated from beliefs and from other subdoxastic states.[31] There are several points to be made here: first, beliefs do not combine with subdoxastic states to yield further beliefs; second, subdoxastic states do not combine with beliefs to yield other subdoxastic states; and third, subdoxastic states do not combine with subdoxastic states in other subsystems to yield further subdoxastic states. Stich sums up the proposal as follows:

> If we think in terms of a cognitive simulation model, the view I am urging is that beliefs form a consciously accessible, inferentially integrated cognitive subsystem. Subdoxastic states occur in a variety of separate, special purpose subsystems. And even when the subdoxastic states within a specialized subsystem generate one another via a process of inference, their inferential interactions with the integrated body of accessible beliefs is severely limited. Similarly, in all likelihood, the potential inferential connections among subdoxastic states in different specialized subsystems are extremely limited or non-existent. (1978: 507-8)

Fodor (1983) draws a closely related distinction between modular and non-modular systems. According to Fodor, modular systems are domain specific and informationally encapsulated in the sense that they take a specific domain of information as input and use a specific domain of information in processing its input.[32] A non-modular system, on the other hand, is domain general and informationally integrated in the sense that it takes inputs from various different modules and processes these inputs in a way that is sensitive to all of the information that is represented in the central system. In Fodor's terminology, the proposal under consideration is that computational processes are modular, whereas cognitive processes are non-modular.

Fodor's distinction between modular and non-modular processes is clearly an important one. However, it is a further question whether this functional distinction provides a basis for explaining the normative distinction between computation and cognition. In order to explain the normative distinction, we would need to invoke a *functional connection thesis* of the following kind:

> The functional connection thesis: an intentional state plays a rational role if and only if it is functionally integrated with other intentional states within a unified nonmodular system.

However, the functional connection thesis is not plausible. On the one hand, not all nonmodular states and processes are cognitive, rather than computational, in the

sense that they are subject to rational assessment. For instance, if perception is cognitively penetrable in the sense that it is influenced by background cognition, then some perceptual processes are nonmodular, but they are not thereby subject to rational assessment.[33] On the other hand, not all modular states and processes are computational, rather than cognitive, in the sense that they are immune from rational assessment. For instance, Spelke (2000) argues that infant cognition is subserved by "core knowledge systems" that are domain specific and informationally encapsulated, while Cosmides and Tooby (1992) argue that adult cognition involves the operation of a series of domain specific and informationally encapsulated modules, such as a "cheater detection" module. Nonetheless, the cognitive processes that are subserved by these specialized modules, including the reasoning involved in detecting cheaters, may be subject to rational assessment.

Similarly, Stich's functional distinction between inferentially integrated and isolated states is an important one, but it cannot explain the normative distinction between beliefs and subdoxastic states. Stich gives the following example to illustrate the functional distinction between beliefs and subdoxastic states:

As another example, suppose that, for some putative rule r, you have come to believe that if r then Chomsky is seriously mistaken. Suppose further that, as it happens, r is in fact among the rules stored by your language processing mechanism. That belief along with the subdoxastic state will not lead to the belief that Chomsky is seriously mistaken. By contrast, if you believe

(perhaps even mistakenly) that r, then the belief that Chomsky is seriously

mistaken is likely to be inferred. (1978: 508-9)

The descriptive point that Stich is making in this example has an obvious normative

counterpart. If I believe that if r, then Chomsky is mistaken, and I also believe that r,

then I am rationally committed to believing that Chomsky is mistaken. By contrast, if

I merely subdoxastically represent that r, then I am not rationally committed to

believing that Chomsky is mistaken. More generally, there is a normative distinction

between beliefs and subdoxastic states such that one's beliefs are subject to rational

assessment in terms of ideals of logical consistency and closure, whereas

subdoxastic states are not subject to the same ideals. But what is the basis of this

normative distinction between beliefs and subdoxastic states?

Broadly speaking, there are two options for explaining the normative

distinction between beliefs and subdoxastic states. On the one hand, we can appeal

to a phenomenal distinction on which the contents of beliefs, unlike subdoxastic

states, are accessible to consciousness as the contents of conscious judgments. On

the other hand, we can appeal to a functional distinction on which beliefs, unlike

subdoxastic states, are inferentially integrated with other beliefs. In order to decide

between these options, we need to consider cases in which the relevant phenomenal

and functional properties are dissociated.

First, let us consider a variation on Stich's example in which the content of

one's representation that r is accessible to consciousness as the content of a

conscious judgment, although it is not functionally integrated with one's other

beliefs and judgments. In that case, one is disposed to judge that r, but one is not disposed to infer that Chomsky is mistaken. This seems rationally defective: if one believes the conditional, and one is disposed to judge the antecedent, then one has a rational commitment to infer the consequent. But if the representation grounds a rational commitment to make an inference, then it is a belief, rather than a subdoxastic state.

Next, let us consider another variation in which the content of one's representation that r is functionally integrated with one's other beliefs and judgments, although it is not accessible to consciousness as the content of a conscious judgment. In that case, one is disposed to infer that Chomsky is mistaken, but one is not disposed to judge that r. Again, this seems rationally defective: if one believes the conditional, but one is not disposed to judge that the antecedent is true, then one has no rational commitment to infer the consequent; indeed, one has a rational commitment not to infer the consequent on those grounds alone. But if the representation does not ground a rational commitment to draw the inference, then it is a subdoxastic state, rather than a belief.

The upshot of this discussion is that it doesn't matter how much we elaborate the functional role of a subdoxastic state to mimic the functional role of belief: it cannot play a rational role unless it is individuated in such a way that its content is accessible to consciousness as the content of a phenomenally conscious judgment. Therefore, I conclude that the functional connection thesis should be rejected and replaced with the rational connection thesis: an intentional state plays a rational role if and only if it is conscious or individuated by its relations to consciousness.

Furthermore, given rational constraints on cognition, the rational connection thesis entails the phenomenal analysis of cognition.

A further question remains. What explains this connection between consciousness and rationality? In my view, the connection is fundamental and so we cannot explain it by deriving it from anything more fundamental in the order of explanation. Nevertheless, I claim that we can acquire some reflective understanding of the connection by recognizing how it explains a connection between rationality and critical reflection. My remarks here will be brief, since I explore these issues in more detail elsewhere.[34]

The concept of rationality is essentially tied to the practice of critical reflection. To a first approximation, a belief is rational if and only if it is based in such a way that it would survive an idealized process of critical reflection. On this conception, the rationality of one's beliefs depends solely upon facts that are accessible to one by means of introspection and a priori reasoning, since these are the methods that constitute the practice of critical reflection.

Given this connection between rationality and critical reflection, we can say that an intentional state plays a role in determining which beliefs it is rational for one to hold only if it is accessible to one by means of introspection. Moreover, it is plausible that an intentional state is accessible to one by means of introspection only if it is either conscious or individuated by its relations to consciousness. Thus, we can argue for the rational connection thesis as follows:

(1) All intentional states that play a rational role are introspectively accessible.

(2) All introspectively accessible states are either conscious or individuated by their relations to consciousness.

(3) So, all intentional states that play a rational role are either conscious or individuated by their relations to consciousness.

This argument provides a theoretical rationale for the rational connection thesis by linking the concept of rationality with the practice of critical reflection. Intentional states that are neither conscious nor individuated by their relations to conscious states do not play a rational role because they are not accessible by means of introspection for use in critical reflection about what to believe. Thus, the connection between rationality and consciousness can be understood by taking into consideration the connection between rationality and critical reflection.

6. Summary and Conclusions

My main goal in this paper has been to argue for a phenomenal analysis of cognition on which all cognitive states are either conscious or individuated by their relations to consciousness. The main argument is that the phenomenal analysis of cognition is indispensable for explaining why cognitive states play a rational role. The general strategy is to explain the connection between consciousness and cognition as a consequence of a more fundamental connection between consciousness and rationality together with rational constraints on cognition.

6.1. Zombies

Can there be a cognitive zombie? Some argue that there cannot be a cognitive

zombie because there cannot be an intentional zombie. In opposition, I have argued

that zombies (like the zombie systems within us) have intentional states and

processes that are indispensable for explaining the relational properties of their

behavior. Nevertheless, their intentional states and processes are not cognitive

states and processes, since they do not play a rational role. In picturesque terms,

zombies are excluded from the space of reasons and so they cannot have cognitive

states and processes, since these are positions and movements within the space of

reasons. In sum, there can be an intentional zombie, but there cannot be a cognitive

zombie, since there cannot be a rational zombie.

6.2. The Unity of the Mental

Do we have any unitary conception of the mental? One historically influential view is

that our concept of the mental is fundamentally disjunctive: all mental states are

either conscious or intentional. Some mental states, such as pains, are conscious but

not intentional, while other mental states, such as beliefs, are intentional but not

conscious. However, there is no further property that mental states such as pains

and beliefs have in common in virtue of which they all count as mental states.[35]

More recently, some philosophers have sought to restore the unity in our

concept of the mental by arguing for tighter connections between consciousness and

intentionality. Some argue for intentionalism, according to which all conscious

states (including pains) are intentional states. Others argue for the intentional

connection thesis, according to which all intentional states (including beliefs) are

either conscious or individuated by their relations to conscious states. If these claims can be sustained, they promise to salvage the unity in our concept of the mental.

However, I have argued that the intentional connection thesis is false: not all intentional states are either conscious or individuated by their relations to consciousness. Moreover, there is a theoretically important distinction between intentional states that can be drawn in terms of consciousness. In particular, cognitive states are distinguished from computational states by the fact that they play a rational role in virtue of their relations to consciousness. If intentionality is the mark of the mental, then the mental is divided between cognition and computation, but cognition is unified by its relations to consciousness.

6.3. Causal and Normative Functionalism

Given a phenomenal analysis of cognition, the prospects for a functional analysis of cognition stand or fall with the prospects for a functional analysis of consciousness. Neither consciousness nor cognition is subject to a reductive style of functional analysis in terms of its causal role. Nevertheless, both consciousness and cognition are subject to a non-reductive style of functional analysis in terms of its normative role. According to a weak version of the rationality constraint on cognition, all cognitive states play a rational role in the sense that they provide reasons or rational requirements to engage in some cognitive processes, rather than others. Moreover, according to the rational connection thesis, all intentional states that play a rational role in this sense are either conscious or individuated by their relations to

consciousness. Thus, consciousness and cognition alike can be analyzed in terms of their normative, rather than their causal, functional roles.

6.4. The Hard Problem of Consciousness

If there is a conceptual connection between consciousness and cognition, then the problem of explaining cognition is intertwined with the problem of explaining consciousness. If consciousness is analyzed in terms of cognition, then the problem of explaining consciousness can be ameliorated by independently solving the problem of explaining cognition. But if cognition is analyzed in terms of consciousness, then the problem of explaining cognition cannot be solved without independently solving the problem of explaining consciousness. So, if the problem of explaining consciousness is a hard problem, then the problem of explaining cognition is a hard problem too.[36]

References

Armstrong, David. 1968. *A Materialist Theory of the Mind*. London: Routledge.

Audi, Robert. 1994. "Dispositional Beliefs and Dispositions to Believe." *Nous* 28 (4): 419-34.

Bermudez, Jose Luis. 1995. "Nonconceptual Content: From Perceptual Experience to Subpersonal Computational States." *Mind and Language* 10 (4): 333-69.

Block, Ned. 1978. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9: 261-325.

Block, Ned. 1990. "Inverted Earth." *Philosophical Perspectives* 4: 53-79.

Block, Ned. 1995. "On a Confusion About a Function of Consciousness." *Behavioral and Brain Sciences* 18: 227-47.

Block, Ned. 2002. "The Harder Problem of Consciousness." *The Journal of Philosophy* 99 (8): 1-35.

Boghossian, Paul. 1989. "The Rule-Following Considerations." *Mind* 98: 507-49.

Bortolotti, Lisa. 2009. *Delusions and Other Irrational Beliefs*. New York: Oxford University Press.

Burge, Tyler. 2010. *Origins of Objectivity*. New York: Oxford University Press.

Campbell, John. 2002. *Reference and Consciousness*. New York: Oxford University Press.

Chalmers, David. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, David. 2003. "The Content and Epistemology of Phenomenal Belief." In *Consciousness: New Philosophical Perspectives*, edited by Q. Smith and A. Jokic. New York: Oxford University Press.

Chalmers, David. 2004. "The Representational Character of Experience." In *The Future for Philosophy*, edited by B. Leiter. New York: Oxford University Press.

Chalmers, David. 2011. "Verbal Disputes." *Philosophical Review* 120 (4): 515-66.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.

Cosmides, Leda and Tooby, John. 1992. "Cognitive Adaptations for Social Exchange." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by J. Barkow, L. Cosmides, and J. Tooby. New York: Oxford University Press.

Davidson, Donald. 1973 "Radical Interpretation." *Dialectica* 27: 313-28.

Davies, Martin. 1987. "Tacit Knowledge and Semantic Theory: Does a Five Percent Difference Matter?" *Mind* 96: 441-62.

Davies, Martin. 1989. "Tacit Knowledge and Subdoxastic States." In *Reflections on Chomsky*, edited by A. George. Oxford: Blackwell.

Davies, Martin. 1995. "Consciousness and the Varieties of Aboutness." In *Philosophy of Psychology: Debates on Psychological Explanation*, edited by C. MacDonald and G. MacDonald. Oxford: Blackwell.

Dennett, Daniel. 1969. *Content and Consciousness*. New York: Routledge.

Dennett, Daniel. 1978. "Toward A Cognitive Theory of Consciousness." *Minnesota Studies in the Philosophy of Science* 9: 201-28.

Dennett, Daniel. 1991. *Consciousness Explained.* Boston: Little, Brown and Co.

Evans, Gareth. 1981. "Semantic Structure and Tacit Knowledge." In *Wittgenstein: To Follow a Rule*, edited by S. Holtzmann and C. Leich. London: Routledge and Kegan Paul.

Fodor, Jerry. 1975. *The Language of Thought.* Cambridge: Harvard University Press.

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

Gendler, Tamar. 2008. "Alief and Belief." *Journal of Philosophy* 105: 634-63.

Gendler, Tamar and Hawthorne, John. 2002. *Conceivability and Possibility*. New York: Oxford University Press.

Horgan, Terry and Graham, George. 2012. "Phenomenal Intentionality and Content Determinacy." In *Prospects for Meaning*, edited by R. Schantz. Berlin: De Gruyter.

Horgan, Terry and Tienson, John. 2002. "The Intentionality of Phenomenology and the Phenomenology of Intentionality." In *Philosophy of Mind: Classical and Contemporary Readings*, edited by D. Chalmers. New York: Oxford University Press.

Kim, Jaegwon. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Koksvik, Ole. 2011. *Intuition*. Ph.D. Dissertation, Australian National University.

Kriegel, Uriah. 2011. *The Sources of Intentionality*. New York: Oxford University Press.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Cambridge: Harvard University Press.

Levine, Joe. 1983. "Materialism and Qualia: the Explanatory Gap." *Pacific*

    Philosophical Quarterly 64: 354-61.

Lewis, David. 1972. "Psychophysical and Theoretical Identifications." *Australasian*

    Journal of Philosophy 50: 249-58.

Lewis, David. 1983. "New Work for a Theory of Universals." *Australasian Journal of*

    Philosophy 61: 343-77.

Loewer, Barry. 1997. "A Guide to Naturalizing Semantics." In *A Companion to the*

    Philosophy of Language, edited by B. Hale and C. Wright. Oxford: Blackwell.

Marr, David. 1982. *Vision: A Computational Investigation into the Human*

    Representation and Processing of Visual Information. New York: Freeman.

McDowell, John. 1994. "The Content of Perceptual Experience." *Philosophical*

    Quarterly 44: 190-205.

McGinn, Colin. 1988. "Consciousness and Content." *Proceedings of the British*

    Academy 74: 219-39.

Milner, David and Goodale, Melvyn. 1995. *The Visual Brain in Action*, New York:

    Oxford University Press.

Nagel, Thomas. 1974. "What is it Like to Be a Bat?" *The Philosophical Review* 83 (4):

    435-450.

Peacocke, Christopher. 1995. "Content, Computation and Externalism." *Philosophical*

    Issues 6: 227-264.

Peacocke, Christopher. 1998. "Conscious Attitudes, Attention, and Self-Knowledge."

    In *Knowing Our Own Minds*, edited by C. Wright, B. Smith, and C. MacDonald.

    New York: Oxford University Press.

Pitt, David. 2004. "The Phenomenology of Cognition, Or, What is it Like to Think that

P?" *Philosophy and Phenomenological Research* 69: 1-36.

Putnam, Hilary. 1981. *Reason, Truth and History*. Cambridge: Cambridge University

Press.

Pylyshyn, Zenon. 1999. "Is Vision Continuous with Cognition? The Case for Cognitive

Impenetrability of Visual Perception." *Behavioral and Brain Sciences* 22: 341-

423.

Quine, Willard van Orman. 1960. *Word and Object*. Cambridge: MIT Press.

Quine, Willard van Orman. 1970. "Methodological Reflections on Current Linguistic

Theory." *Synthese* 21: 386-398.

Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton

University Press.

Ryle, Gilbert. 1949. *The Concept of Mind*. Chicago: Chicago University Press.

Schwitzgebel, Eric. 2002. "A Phenomenal, Dispositional Account of Belief." *Nous* 36:

249-75.

Schwitzgebel, Eric. 2010. "Acting Contrary to Our Professed Beliefs, or the Gulf

Between Occurrent Judgment and Dispositional Belief." *Pacific Philosophical

Quarterly* 91: 531-553.

Searle, John. 1990. "Consciousness, Explanatory Inversion and Cognitive Science."

Behavioral and Brain Sciences 13: 585-96.

Searle, John. 1992. *The Rediscovery of Mind*. Cambridge: MIT Press.

Shoemaker, Sydney. 1975. "Functionalism and Qualia." *Philosophical Studies* 27:

291-315.

Siewert, Charles. 1998. *The Significance of Consciousness.* Princeton: Princeton

    University Press.

Smithies, Declan. 2011. "What is the Role of Consciousness in Demonstrative

    Thought?" *The Journal of Philosophy* 108 (1): 5-34.

Smithies, Declan. 2012a. "Introspection and Consciousness." In *Introspection and*

    Consciousness, edited by D. Smithies and D. Stoljar. New York: Oxford University

    Press.

Smithies, Declan. 2012b. "Moore's Paradox and the Accessibility of Justification."

    Philosophy and Phenomenological Research 85 (2): 273-300.

Smithies, Declan. Forthcoming a. "The Phenomenal Basis of Epistemic Justification."

    In *New Waves in Philosophy of Mind*, edited by J. Kallestrup and M. Sprevak.

    Basingstoke: Palgrave Macmillan.

Smithies, Declan. Forthcoming b. "The Nature of Cognitive Phenomenology."

    Philosophy Compass.

Smithies, Declan. Forthcoming c. "Why Care About Justification?" In *Epistemic*

    Evaluation: Point and Purpose in Epistemology, edited by J. Greco and D.

    Henderson. New York: Oxford University Press.

Spelke, Elizabeth. 2000. "Core Knowledge." *American Psychologist* 55: 1233-43.

Stich, Stephen. 1978. "Beliefs and Subdoxastic States." *Philosophy of Science* 45: 499-

    518.

Stich, Stephen, and Warfield, Ted. 1994. *Mental Representation: A Reader*. Oxford:

    Blackwell.

Strawson, Galen. 1994. *Mental Reality*. Cambridge: MIT Press.

Strawson, Galen. 2004. "Real Intentionality." *Phenomenology and the Cognitive Sciences* 3: 287-313.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Zimmerman, Aaron. 2007. "The Nature of Belief." *Journal of Consciousness Studies* 14: 61-82.

[1] See Chalmers (1996) for an influential discussion of zombies in the context of metaphysical debates about the nature of consciousness.

[2] See the introduction and essays in Gendler and Hawthorne (2002) for more on the distinctions between physical, metaphysical, and conceptual possibility. To a first approximation, it is conceptually possible that p if and only if it is not a priori that p is false. For present purposes, I assume that metaphysical possibility entails conceptual possibility, but not vice versa.

[3] Bifurcationist themes can be found in the work of Block (1978), Chalmers (1996), and Kim (2005), although in more recent work, Chalmers (2003) has moved away from bifurcationism.

[4] The project of giving a conceptual analysis of consciousness in terms of its role in cognition is closely associated with proponents of analytical functionalism, such as Armstrong (1968), Lewis (1972) and Shoemaker (1975). Others, including Dennett (1978, 1991), endorse cognitive theories of consciousness that are proposed as empirical theories, rather than as conceptual analysis.

[5] Proponents include McGinn (1988), Searle (1990, 1992), Strawson (1994, 2004), Davies (1995), Campbell (2002), Chalmers (2003), Kriegel (2011), Smithies (2011), Horgan and Graham (2012).

[6] Compare Block's (1995) distinction between the phenomenal concept of consciousness and various functional concepts of consciousness, including access consciousness and monitoring consciousness; see also Chalmers (1996: 25-31) for a related distinction. Both Block (2002) and Chalmers (2003) endorse *phenomenal realism* – that is, the thesis that the phenomenal concept of consciousness cannot be defined or analyzed in terms of any functional concept.

[7] See Chalmers (1996: xi-xiii) on the hard problem of consciousness. See also Nagel (1974) and Levine (1983) for two influential discussions of the explanatory gap.

[8] Thus, Chalmers (1996: 172) writes, "Cognition can be explained functionally; consciousness resists such explanation." See also Block (1990: 53-4) and Kim (2005: 161-70).

[9] Smithies (forthcoming a) provides an overview of my position on the normative role of consciousness, while Smithies (2011, 2012a) explores applications to demonstrative concepts and introspection.

[10] Ryle (1949) and Wittgenstein (1953) are often interpreted as proponents of analytical behaviorism, although this interpretation is controversial. Analytical functionalism has many influential proponents, including Armstrong (1968), Lewis (1972) and Shoemaker (1975). Others, including Fodor (1975), endorse functionalism as an empirical theory, rather than a conceptual analysis, of cognition.

[11] Thus Block (1990: 53-4) writes, "The 'containment response'…would be to give up on functionalism as a theory of experience (or at least of its qualitative aspect), retaining functionalism as a theory of the cognitive aspect of the mind."

[12] Block (1990: 54) acknowledges that "the containment response…arguably commits its proponents to the possibility of a 'zombie', a being that is like us in cognitive states but totally lacking in qualia."

[13] These considerations loom large in Ryle's (1949: Ch. 1) discussion of Descartes' Myth.

[14] This dispositional analysis of belief is consistent with Audi's (1994) distinction between dispositional beliefs and mere dispositions to believe, since we can distinguish between first-order dispositions towards judgment and second-order dispositions to acquire those dispositions.

[15] For detailed discussion and defence of this assumption, see Smithies (forthcoming b). See also Strawson (1994), Siewert (1998), Horgan and Tienson (2002), and Pitt (2004). The terminology of 'content-specific' and 'attitude-specific' phenomenal character is borrowed from Ole Koksvik (2011); see also Horgan and Tienson (2002) for a related distinction between the phenomenology of intentional content and the phenomenology of attitude-type.

[16] See Horgan and Tienson (2002) and Chalmers (2004) for related proposals.

[17] See Zimmerman (2007) and Gendler (2008) for additional cases of absentmindedness, phobia, and prejudice, in which a subject is disposed to behave as if she believes a proposition, although it seems

wrong to say that she believes it. According to Gendler, the subject's behavioral dispositions do not reflect her beliefs, but rather a distinctively practical attitude, which she dubs 'alief'.

[18] Strawson (1994, 2004) endorses the strong version of the intentional connection thesis, while Searle (1990, 1992), Kriegel (2011) and Horgan and Graham (2012) endorse weaker versions.

[19] Here, I set aside Kriegel's (2011) interpretationist theory on which unconscious intentionality depends on a conscious act of interpretation.

[20] See Horgan and Graham (2012) for the most sustained development of this line of argument.

[21] See the essays in Stich and Warfield (1994) for a representative sample of causal theories of intentionality. See Quine (1960), Davidson (1973), Putnam (1981), and Kripke (1982) for classic discussions of the problem of radical indeterminacy. See Boghossian (1989) and Loewer (1997) for discussion of these problems in connection with causal theories of intentionality.

[22] See Boghossian (1989) for sympathetic discussion of anti-reductionism about mental content.

[23] Searle (1990) and McDowell (1994) endorse instrumentalism about the role of content in computational explanation, whereas Strawson (2008) endorses a more extreme form of eliminativism about all unconscious intentionality. See Peacocke (1995) for a defence of the claim that content plays an indispensable role in computational explanation.

[24] Stich (1978: 499) defines subdoxastic states as "psychological states that play a role in the proximate causal history of beliefs, though they are not beliefs themselves." For current purposes, we can assume that subdoxastic states are intentional states, but not phenomenally conscious states.

[25] See Evans (1981) and Davies (1987) for this account of tacit knowledge.

[26] Compare Burge's (2010: 374-6) discussion of unconscious perception.

[27] See Bortolotti (2009) for an overview and discussion of the empirical evidence on irrational cognition.

[28] See Peacocke (1995) for arguments that we need to recognize a semantic conception of computation in addition to a purely syntactic conception of computation.

[29] See Smithies (2012b, forthcoming c) for arguments that epistemic rationality or justification is accessible in the sense that one has justification to believe a proposition if and only if one has

justification to believe that one does and, equally, one lacks justification to believe a proposition if and only if one has justification to believe that one does.

[30] See Smithies (2011, 2012a) for discussion of attempts to undercut the rational role of consciousness by appealing to functional properties of accessibility or metacognition. Here, I focus instead on the functional property of inferential integration, which I take to be closely related to the functional property of systematicity, since a representational system is inferentially integrated only if it is systematic. I plan to discuss systematicity in more detail elsewhere.

[31] Stich (1978: 517) claims that beliefs are distinguished from subdoxastic states not only by their inferential integration, but also by their accessibility to consciousness, although he does not explain why these properties come together or argue that they cannot be dissociated.

[32] Fodor (1983: 71) claims that informational encapsulation is "the essence of…modularity", although he also claims that modules typically exhibit a cluster of related symptoms, including inaccessibility to consciousness, fast and mandatory operation, shallow outputs, fixed neural architecture and characteristic patterns of breakdown and development.

[33] Pylyshyn (1999) argues that early vision is cognitively impenetrable, but this view is consistent with the claim that higher-level vision is cognitively penetrated.

[34] See Smithies (2012a, Forthcoming a & c).

[35] Rorty (1979: 22) gives a very clear statement of this view.

[36] More or less distant ancestors of this paper were presented at the CUNY Graduate Center in 2005, the Australian National University in 2007, the Australasian Association of Philosophy in 2009, my epistemology seminar at the Ohio State University in 2009, the Columbia-Barnard Perception Workshop in 2012 and Oberlin College in 2012. Many thanks to audiences on all those occasions and especially to Ned Block, Paul Boghossian, David Chalmers, Benj Hellie, Uriah Kriegel, Geoff Lee, Christopher Peacocke, Abe Roth, Richard Samuels, Eric Schwitzgebel, Nico Silins, and Jeremy Weiss.