Original Paper UDC 165.24: 159.923:2
Received September 5th, 2007

Julian Kiverstein

University of Edinburgh, Department of Philosophy, George Square, GB-Edinburgh EH8 9JX
j.kiverstein@ed.ac.uk

Consciousness, the Minimal Self, and Brain[1]

Abstract

This paper explores the possibility of a neuroscientific explanation of consciousness, and what such an explanation might look like. More specifically, I will be concerned with the claim that for any given experience there is neural representational system that constitutes the minimal supervenience base of that experience. I will call this hypothesis the minimal supervenience thesis. I argue that the minimal supervenience thesis is subject to two readings, which I call the localist and holist readings. Localist theories seek to identify the minimal supervenience base for specific experiences. They sideline questions about the nature of creature consciousness, treating the neural basis of creature consciousness as merely a causally necessary background condition for a particular conscious experience. Holists on the other hand prioritise creature consciousness and argue that we can only account for particular states of consciousness in the context of an account of creature consciousness. I argue that any scientific explanation of consciousness must account for what I will call a minimal sense of self that is intrinsic to every conscious state. Holist theories are best able to accommodate this feature. I end by arguing that the Dynamic Sensorimotor (DSM) account of consciousness can be combined with a holist account of the neural basis of consciousness. Such a combination of views corrects for the opposition to the minimal supervenience thesis found in some prominent defenders of the DSM account (e.g. Alva Noë and Evan Thompson). It also provides a framework for developing a neuroscientific account of the minimal sense of self.

Keywords

neural correlates of consciousness, state consciousness, Creature Consciousness, minimal sense of self, pre-reflective self-awareness, sensorimotor dynamics

1. What is the problem of consciousness?

Discussions of the problem of consciousness typically begin with the observation that conscious experiences are essentially *subjective*. There is something it is like for a subject to see a ripe bulgy tomato, to touch silk, to smell fresh coffee, to hear the sound of a passing siren, to fall in love – to mention just a selection of the almost infinite variety of conscious episodes that populate our mental lives from moment to moment. The problem of consciousness arises when we attempt to understand how each of these experiences could be the outcome of the brain's information processing. Cognitive neuroscience is amassing a growing body of knowledge about how our brains process in-

1
Work on this paper was funded by the AHRC under the ESF Eurocores Consciousness in the Natural and Cultural Context scheme for the CONTACT (Consciousness in Interaction) Project, AH/E511139/1. Many thanks to Susan Hurley, Andy Clark and Till Vierkant for helpful discussion of some of the ideas in this paper, and to Zdravko Radman for inviting this contribution.

formation. Yet if asked, many neuroscientists will admit to having very little idea as to how conscious experience could be the product of this information processing. Two leading neurobiologists, Francis Crick and Christof Koch (2003) claim:

"No one has produced any plausible explanation as to how the experience of the redness of red could arise from the actions of the brain. It appears fruitless to approach this problem head-on." (p.119)

In a similar spirit Nancy Kanwisher (2001), in an influential paper on the neural correlates of perceptual awarness, has written:

"I hope in this article to show that scientific evidence can bear importantly on a number of questions about the nature of perceptual awareness. However, it probably cannot answer all such questions. In particular, I will not tackle the question of why perceptual awareness feels like anything at all because it is not clear that even a rich understanding of the cognitive and neural events that constitute perceptual awareness will provide any clues about how to answer it." (p. 90–1)

The problem of consciousness arises when we attempt to integrate two types of knowledge: the third-person knowledge we have of the brain and the first-person knowledge by acquaintance that we have with our own experiences as we live through them. Each of us is acquainted directly with his or her conscious experiences, and this acquaintance makes possible a certain kind of first-person knowledge. This knowledge is *first person* knowledge because the objects of this type of knowledge are one's own conscious mental states as they are represented from a first person point of view. Neuroscience, on the other hand, supplies third-person descriptions of neural processing. The knowledge it supplies of the brain is *third-person* knowledge. There is a compelling intuition that no matter how deep our third-person knowledge of the brain reaches, there will remain some facts about conscious experience that we cannot learn from third-person knowledge of the brain. We cannot learn why there should be something it is like to be conscious, or why our experiences have the particular subjective qualitative character they do. There is a gap separating our first-person knowledge of experience and our third-person knowledge of the brain. Our third-person knowledge of the brain doesn't necessitate or conceptually entail what we know of our own experiences from a first-person point of view. What we know of our own experiences from a first-person point of view cannot be deduced from any third-person descriptions of brain processing. I will henceforth label this intuition the *gap intuition.*

The gap intuition has been taken to support the metaphysical thesis that conscious properties and neural properties are numerically distinct kinds of properties. (By "conscious properties" I mean properties like being conscious rather than unconscious or having an experience with a particular subjective character.) Suppose for the sake of argument that every situation that is coherently conceivable is also metaphysically possible.[2] There may be some situations that we think we can coherently conceive of but that are not metaphysically possible. We can say of such cases that we were mistaken about the situation we were conceiving of: we thought we were conceiving of a situation that is metaphysically impossible, but in fact we were conceiving of a distinct metaphysically possible situation. We do not have first-person authority about what it is we are conceiving. We have said that third-person knowledge of the brain doesn't conceptually entail first-person knowledge of experience. (This is just a statement of the gap intuition.) To accept the gap intuition is to say that we can conceive of everything we know about the brain obtaining in

the absence of what we know about experience from the first-person point of view. Suppose this is coherently conceivable and that whatever is coherently conceivable is also metaphysically possible. Then it follows that it is metaphysically possible for neural properties to be instantiated without conscious properties also being instantiated. If conscious properties and neural properties were identical, it would not be possible for one type of property to be instantiated without the other. Hence we can conclude that conscious properties and neural properties are not identical, which is to say that neural properties and conscious properties are distinct types of property.[3]

Materialist philosophers must of course reject such a conclusion, and I will follow them in doing so. Materialists have tended to respond to the gap intuition in two broad ways. Some have admitted that we do not yet have detailed knowledge as to how the brain could generate conscious experience, but they claim that this gap in our knowledge is one that will in time be remedied.[4] As our knowledge of the brain grows so also will our understanding of the relationship between consciousness and the brain. Of course any materialist will owe an explanation as to why so many philosophers have been tempted to draw a metaphysical conclusion from the gap intuition. One possible diagnosis of this error would be to argue that it rests on the mistaken belief that what we know of our experiences from the first-person perspective must be conceptually entailed or necessitated by our third-person knowledge of the brain (see, e.g. Block & Stalnaker 1999). When we find no such entailment we are tempted to conclude that neural and conscious properties are distinct kinds of properties. However why should we accept that first-person knowledge must be conceptually entailed by third-person knowledge of the brain? If we reject such a requirement we no longer have grounds for drawing any anti-materialist metaphysical conclusions.

Other materialist philosophers have been more willing to concede something to the gap intuition.[5] They accept that there is a gap separating what we know of our experiences from a first-person perspective and what we know of the brain, but because they are materialists they deny that the gap is metaphysical. They deny that conscious mental states have properties distinct from the properties of brains and brain processes. We find the gap intuition compelling, these philosophers argue, because of the different concepts we deploy when we think about the brain and when we think about consciousness. When we conceive of neural processes we do so using what we can call "neural con-

2
For discussion of the relation between conceivability and possibility see the essays in Gendler & Hawthorne (2002), and for a defence of this supposition the chapter by Chalmers (2002a) in this collection. For arguments challenging this supposition see Balog (2000); Hill & McLaughlin (1999); Block and Stalnaker (1999) and Papineau (2002).

3
This argument is a simplification of one to be found in Chalmers (1996).

4
Philosophers in this first camp take a similar stance to Chalmers' (2002b) type A materialists. Chalmers offers as examples of type A materialists Armstrong (1968); Dennett (1991); Dretske (1995); Harman (1990); Rey (1995) and Ryle (1949). There is however a key difference in the position I am articulating and Chalmer's type A materialists. Chalmers characterises the latter as either analytic functionalists or as logical behaviourists whereas the neurophilosophical approach I will pursue in this paper is perhaps closer to what McLaughlin (2007) calls "type materialism". Type materialists hold that the reason why a neurobiological state N is nomologically correlated with a type of experience E is because N = E.

5
Chalmers (2002) calls philosophers in this second camp "type B materialists". Examples are Loar (1997); Sturgeon (2001); Papineau (2002); Perry (2001); Tye (2000, ch. 2).

cepts", and when we introspect and think about what it is like to undergo a particular experience we do so by deploying "phenomenal concepts". To acquire mastery of a phenomenal concept one must have undergone an experience of the type the phenomenal concept denotes. This is not the case for the neural concepts we employ in our thinking about the brain. It follows that one might very well be capable of thinking about the neural activity underlying a particular experience, though one could not think about the experience itself under the relevant phenomenal concept because one had not had an experience of the type the phenomenal concept denoted. The gap intuition is right then to claim that there is no conceptual entailment from third-person knowledge of the brain to first-person knowledge of experience. However this is a consequence of what it takes to master a phenomenal concept and tells us nothing about the nature of conscious experience.[6]

In this paper I will take up a response to the gap intuition that belongs (with some important qualifications) to the first of the camps just described. I will begin by exploring the claim that for any given experience *e* there exists a neural system – a configuration of neural activity – that constitutes the minimal supervenience base for the occurrence of *e*. Call this "the minimal supervenience claim".[7] In §2 I will explain how this claim is subject to two very different readings, which I will call the *localist* and *holist* readings respectively. Localist theories are distinguished from holist theories in treating state consciousness as a separate problem from creature consciousness. (The distinction between "state" and "creature" consciousness originates with Rosenthal 1990/1997 & 1993/2005). In §3 I will argue that the localist reading is at best incomplete to the extent that it sidelines questions about creature consciousness. Creature consciousness, I will suggest, is best understood as conferring on a creature a minimal sense of self, which is an intrinsic or core feature of every conscious experience. To the extent that localist theories ignore creature consciousness, they fail to explain this core feature of experience. In §4 I argue that holist theories are to be preferred on the grounds that they can provide an account of this minimal sense of self. Indeed I will sketch two such accounts. I finish by arguing that a holist account of the neural basis of consciousness can be combined with a dynamic sensorimotor account (henceforth DSM) of consciousness in a way that promises to answer the questions the gap intuition presses. DSM accounts have previously positioned themselves as opposed to the minimal supervenience claim (see for instance Thompson & Noë (2004); Noë (2004, ch.6; 2006 & 2007); Cosmelli & Thompson (2007)). I argue that this opposition is best understood as being targeted at a localist conception of neural correlates. A holist understanding of neural correlates is perfectly consistent with a sensorimotor account of consciousness.

Before we get down to business, let me say something brief about the second type of materialist response to the gap intuition. Materialists belonging to this second camp take the gap intuition to be a consequence of certain peculiarities specific to phenomenal concepts, the concepts we employ in our thinking about conscious experience. They seem to suppose that what motivates the gap – the subjective character of our experience – is really just a consequence of the peculiar concepts we bring to bear in thinking about our experience. Understand how these concepts work and, they say, we will have explained what strikes many as missing from neural information processing explanations of consciousness.

Certainly, there are important and interesting connections between the subjectivity of experience and the concepts we bring to bear in thinking about experience, but understanding these connections won't be sufficient to close

the explanatory gap. Experiences have a subjective content and character that doesn't derive from the concepts we bring to bear in thinking about our experiences. We can see this just by reflecting on the case of animals and infants who most likely do not engage in much, if any, thinking about their experiences. Yet still there is something it is like for an animal or infant to undergo a conscious experience: animals and infants have a subjective mental life. A theory of phenomenal concepts will inform us about the kind of subjectivity that accompanies what we might call "reflective consciousness" – the sort of consciousness that goes along with reflective or introspective thought about one's own experience, or when we conjure up an experience in imagination. However it will do nothing to explain we might call "pre-reflective self-consciousness", the sort of acquaintance we have with our conscious states prior to any act of reflection. Yet as Zahavi (2006: 116–124) explains, there is something it is like to undergo an experience because pre-reflective self-consciousness forms a part of the structure of our experiences. Pre-reflective self-consciousness confers on experience what Zahavi describes as a "first-person givenness". I have a sense that the objects I perceive are ones that are perceived from my perspective. I also have a sense that the thoughts I think are my thoughts. All of my conscious states of mind are given to me under a first-person mode of presentation. It is this first-person givenness that I referred to earlier in this section when I said that conscious subjects are *acquainted* with their experiences. Moreover, it is this first-person givenness that must be explained if we are to account for the subjective character of experience. Yet materialist theories that appeal to phenomenal concepts do nothing to explain this defining feature of subjective experience. They are at best able to explain the kind of subjectivity that accompanies reflective consciousness. If a materialist response to the gap intuition is to be had, it won't come from attempts at deflating the problem by providing a theory of phenomenal concepts. Rather a science of consciousness must confront the problem head-on and show how the brain's information processing might realise an organism's acquaintance with its own experiences. What follows in this paper is intended to show how a science of consciousness is at least addressing this problem.

2. Conceptual preliminaries

David Rosenthal (1990/1997 & 1993/2005) has made a distinction between "state-consciousness" (i.e. states there is something it is like for a creature to

6
I don't mean to suggest that these two responses are mutually exclusive. Papineau (2002) arguably combines both responses in his defence of materialism.

7
Why have I chosen to pursue a neuroscientific approach to explaining consciousness as opposed to a functionalist approach like the one found in Dennett (1991)? Does this commit me to a version of reductive physicalism? Very briefly, I doubt that the kinds of explanations offered by neuroscientists are ultimately that different from functionalist explanations. Neuroscientific explanations commonly individuate neural states in terms of their function, and the contribution they make to some information-processing task. Thus I would argue that neuroscientific explanations are best understood as a species of functional explanation. On the question of reductive physicalism, this raises complex metaphysical issues I do not have the space to grapple with. I commit myself to at most a naturalism that seeks to show how phenomenological descriptions of experience cohere with scientific explanations of consciousness. I do take a stand on metaphysical questions relating to the truth or otherwise of physicalism. Questions of this kind can only be settled by working within our existing scientific theories, but there is little reason to conclude from these theories that physics can give us an exhaustive description of what there is. However, to fully defend such an assertion takes me well outside of the issues I am discussing in this paper.

be in) and "creature-consciousness", which he characterises as the property of being awake. Rosenthal takes state consciousness to be the problematic concept; he writes:

> "It is the notion of a mental state's being conscious that occasions such difficulty in understanding what consciousness amounts to…no special problems impede our understanding of what it is for a creature to be a conscious creature. A creature's being conscious means that it is awake and mentally responsive. Being awake is presumably an unproblematic biological notion." (1993/2005: 46)

In a similar fashion Chalmers (2000) distinguishes explanations of the contents of consciousness from explanations of background consciousness. He characterises explanations of the contents of consciousness as concerned with specific conscious states which he contrasts with overall states of consciousness, like "being awake, being asleep, dreaming, being under hypnosis…" (op. cit., 18).

I agree with Bayne (2007) and Hohwy (2007 & in progress) that we need to take some care in interpreting this distinction. In particular we should not conceive of state and creature/background consciousness as distinct properties of an organism. We might naturally think of state consciousness as a property of some mental states and creature consciousness as a property of individual organisms under certain conditions. To explain the neural basis of state consciousness we must look for the neural correlates of particular experiences, and of the individual features these experiences represent. We might suppose that this is a distinct project from explaining the neural basis of the contrast between consciousness and its absence. I will call theories that take the neural correlates of state consciousness to be distinct from the neural correlates of creature consciousness, *localist* theories. They are localist in the sense that they attempt to pinpoint or localise the neural activity that is correlated with specific types of experience. It should be noted that localisation of this kind need not be read as the view that there are physically discrete areas of the brain that encode particular contents, in a fashion akin to so-called "grandmother cells". A localist theory could accept that experience is correlated with large-scale distributed neural activity spread across geographically disparate areas of the brain. Such a theory would owe an explanation of how this activity is integrated and coordinated. Providing such an account can be given that does not also explain creature consciousness, the theory will qualify as a localist theory. The defining characteristic of a localist theory (as we will see in §3) is that (1) it takes specific experiences to be correlated with localisable neural activity, and (2) it takes state consciousness to be something that can be investigated independently from creature consciousness. Localists are interested only in identifying the neural basis of particular sorts of experience, and bracket the question of what it is in the brain that makes the difference between the presence and absence of consciousness. The assumption that state consciousness can be investigated independently of creature consciousness is rejected by theories that I will label *holist*. Holist theories study how the neural correlates of creature consciousness and the neural correlates of state consciousness interact. I call these theories *holist* because they take the key to explaining the neural basis of consciousness to be the *integration* and coordination of neural activity in widely separated areas of the brain.

Before we turn to the debate between localists and holists we must make more precise the notions of creature consciousness and of the neural correlates of consciousness. On the question of how to understand "creature consciousness", again I am in complete agreement with Bayne (2007: fn1, p.18; also

see Hohwy in progress) that it can't be right to define creature consciousness in terms of being awake, as Rosenthal has suggested (see for instance the above quote). Patients in vegetative states are "awake" in the sense that their eyes open as part of the normal sleep-wake cycle, but they are not creature-conscious. They undergo states of "unconscious wakefulness" (Merker, 2007: 112; also see, Damasio (1999: ch.3) & Laureys (2005)). The same is true of patients in the midst of an absent epileptic seizure: they are unconscious – this is what makes the seizure absent – but they are nevertheless awake. Epileptics can be in the midst of a conversation when a seizure strikes, stopping them in their tracks sometimes in mid-sentence. Once the seizure is over patients will often continue where they left off with no recollection as to what had just happened (Damasio 1999: 95–101). Conversely, subjects can be asleep but consciously dreaming. These subjects are surely enjoying something in the way of creature consciousness, despite not being awake.

The notion of creature consciousness is supposed to pick out a property that marks the contrast between the presence of consciousness and its absence. I suggest that what will make the relevant difference is whether the creature is acquainted with its own states of mind under a first-person mode of presentation. The difference between conscious and non-conscious states of mind is that conscious states have, as a core or structural feature, what I earlier called "pre-reflective self-consciousness". (Notice I am talking about *state*-consciousness here, and not creature consciousness yet.) The notion of "pre-reflective self-consciousness" is intended to pick out a core, structural feature of consciousness, *structural* in the sense that it is an essential and intrinsic feature of every conscious state of mind. Pre-reflective self-consciousness is intrinsic to those mental states a creature is acquainted with under a first-person mode of presentation.

How should we understand this mode of acquaintance? The objects of my experience are given to me in a unique and distinctive way, even when you and I both undergo experiences of one and the same object. Say you and I are both looking at the last slice of cake on the plate. We are both aware of one and the same slice of cake, and we are both aware of the cake in the same type of way, under a first-person mode of presentation. However, what you can have no awareness of is the first-person givenness of my experience. Similarly what I have no awareness of is the first-person givenness of your experience. States that are given in this way have built into them a *minimal sense of self* – my conscious states immediately reveal themselves as mine. Experiences that I am acquainted with under a first-person mode of presentation are ones that I am aware of as my own. You cannot be aware of anyone else's experience but your own under a first-person mode of presentation. To be aware of a state of mind *as* your own is to have a minimal sense of self. I suggest that it is this minimal or core sense of self that marks the difference between the presence of consciousness and its absence. When consciousness is present so also will this minimal sense of self be present. Henceforth, when I talk of creature consciousness we can take this concept to be co-extensive with the minimal sense of self that forms an intrinsic part of every conscious experience. I will be defending holism by arguing that the project of explaining state consciousness is bound up with the project of explaining this minimal sense of self. This must be the case, since this minimal sense of self is an intrinsic part of every conscious state.

What is meant by talk of the *neural correlates* of consciousness (henceforth "NCCs")? Ned Block (2005) defines a neural correlate of *phenomenal* con-

sciousness[8] as "the minimal neural basis of the phenomenal content of an experience" (p. 46). While in Block (forthcoming) he talks of the "core neural basis of experience", David Chalmers (2000) offers a similar definition; according to Chalmers:

> "A NCC (for content) is a minimal neural representational system N such that representation of content in N is sufficient, under conditions C, for a representation of that content in consciousness." (p. 31)

To say that a neural representational system N is *sufficient* for the occurrence of an experience is to say that nothing else is required in order for the experience to occur other than activity in the population of neurons of which N is composed. To make the same point in positive terms, the claim is that as a matter of nomological necessity if a brain instantiates a neural representational system N then a subject will enjoy an experience of type E. Block and Chalmers talk of *minimal sufficiency* to demarcate a system whose elements include only those cells whose excitation or inhibition is required for the occurrence of an experience with a given content. Many other cells will most likely be active when an experience occurs even though their activation is entirely unrelated to the occurrence of the experience in question. These cells will not form a part of the representational system that constitutes the NCC for a given experience because they do not form a part of the minimal neural representational system required for a given experience.

Block and Chalmers are both careful to say that it is only under certain conditions that activity in N will suffice for a given experience. Both follow Shoemaker (1981) in making a distinction between a core realiser and a total realiser. Block defines a *core realiser* as the part of a neural representational system that distinguishes one conscious content from another (forthcoming, ms: p. 3). Thus consider a visual experience as of motion. MT/V5 is a strong candidate for the core realiser for such an experience. Damage to this region causes akinetopsia or motion-blindness, thus activity in this area would seem to be necessary for visual experience as of motion. However it may not be sufficient: in addition recurrent feedback between MT/V5 and V1 may also be required (Block 2005: 46; also see Lamme 2004; Pascual-Leone & Walsh 2002).

Block and Chalmers both think of core realisers along the lines of Mackie's INUS condition (Mackie 1974) – they are an insufficient but necessary part of an unnecessary but sufficient condition for an experience with a particular type of content. The cells that make up a core realiser could not realise a conscious experience apart from wider activity in the brain – in this sense they are insufficient. To suppose they could is to suppose we could cut cells from a brain and place them in a bottle and they would continue to support a conscious experience (c.f. Block forthcoming, ms: p. 2). Yet this is obviously something we cannot do. Only given the right background conditions will a core realiser prove sufficient for a particular conscious experience.

The total realiser for a given experience will include the cells that make up a core realiser plus the background conditions required for these cells to play the role of realising an experience. Those cells that do not form part of the core realiser will constitute the background conditions that must be in place if the core realiser is to do the work of supporting a given type of experience. Block (forthcoming, ms: p. 3–4) makes a further distinction between *causally necessary* background conditions such as cerebral blood flow, and what he calls *constitutive* background conditions. He gives, as an example of the lat-

ter, the activation of the upper brainstem. The upper brainstem does not form a part of a core realiser for a given experience because it doesn't play a role in explaining the contents of consciousness, but it may nevertheless make a constitutive contribution to realising experience, as has been argued by Merker (2007) and Parvizi & Damasio (2001). I will have more to say about the notion of constitutive background conditions later in §3, where I will argue that it is difficult to make sense of the distinction between constitutive background conditions and the core realiser of a given state.[9]

How do we determine which neurons belong to a particular neural representational system and which do not? How do we know when we have the same neural representational system, and when we have a different neural representational system? These are just the sorts of question we must settle if we are to establish that a putative neural representational system is a minimal system, in the sense described above. Localists take the part of a neural representational system that is the core-realiser to be the NCC, and treat the non-core part of a total realiser as part of the background or enabling conditions. The localist's core realiser is the correlate of an experience with a particular content. It is the minimal sufficient condition for an experience of a particular type, but only in the context of the right causal background conditions. Amongst these background conditions will be neural activity that is correlated with creature-consciousness. Creature consciousness is taken to be at best a causally necessary background condition. Holists, on the other hand, deny that creature consciousness is just background enabling conditions. They argue that neural activity correlated with creature consciousness should also be treated as making a constitutive contribution to realising an experience. Thus, holists challenge the localist conception of the distinction between the core and the background conditions for a given experience. I give these theories the label *holist* because they emphasise the role of *integration* in coordinating the neural activity underling particular states of consciousness, and the neural activity that forms the basis for creature consciousness. In the next section I outline in more detail, and argue for, the incompleteness of localist theories of NCCs.

3. The localist account of NCCs

Localists offer hypotheses about the neural basis of the contents of specific conscious experiences. The descriptions they offer of consciousness are therefore neural level descriptions as contrasted with functional or computational level, and personal-level or phenomenological descriptions. However, it should be noted that localists also tend to subscribe to particular personal-level or phenomenological description of consciousness. They tacitly assume what Searle (2000) has called the "building block model of consciousness". They conceive, for instance, of the contents of *visual* consciousness as composed of features such as colour, shape, size, volume, orientation *etc*. Any given visual experience is built out of these features. Notice that this is a

8
Block (2005) argues that the neural correlates for phenomenal and access consciousness may well be distinct kinds of neural activity. Again I will remain neutral on the phenomenal/access consciousness distinction. In what follows it can be assumed that when I talk of "subjective experience" I mean phenomenal conscious experience – experience that there is something it is like to undergo.

9
Hohwy (in progress) also makes this point in a footnote discussing different concepts of consciousness.

claim about what is consciously experienced – as such it is a personal or phenomenological level claim. However this phenomenological assumption informs and shapes how the localist goes about investigating the neural basis of consciousness. To the extent that the localist subscribes to the building block model s/he is likely to think of a subject's consciousness at any given moment as made up of a multiplicity of different experience, each with its own neural correlate. These contents will no doubt often overlap. Some will relate to one and the same object. Others will be of features belonging to distinct objects that occupy a common space. However, insofar as the building block model decomposes the contents of consciousness into distinct elements, it will not be looking for a structural feature that is common to all of these various contents. I will argue, at the end of this section, that this omission is fatal to the building block model, but also to localist theories to the extent that they assume such a model. However, before I can mount such an argument, we must have before us some examples of localist theories.

Semir Zeki (2003; 2007, also see Bartels & Zeki 1998) is perhaps the most radical advocate of localism. Zeki has argued for the existence of what he calls "micro-consciousnesses". He takes as his starting point the idea that geographically distinct areas in visual regions of the brain have the function of detecting different visual attributes. He claims, for instance, that areas V4 and V5 have the function of processing information about colour and motion respectively. On the basis of clinical data, Zeki argues for a double dissociation of processing in these areas. Patients can suffer damage to V5 resulting in *cerebral akinetopsia* (the inability to visually perceive motion) but so long as V4 is spared they will have a normal ability to visually perceive colour. Contrariwise, patients can suffer damage to V4 resulting in *achromotopsia* (an inability to see the world in colour) but if area V5 is intact in these patients they will have an unimpaired ability to see moving objects. Zeki takes this double dissociation to support the claim that V4 and V5 are functionally specialised areas dedicated to processing information about colour and motion respectively. He goes on to argue that processing in these areas may also be correlated with visual *experience* of colour and motion respectively. Zeki suggests that what determines whether a subject visually experiences a given stimulus or not is "the strength of activation" in a given specialised area. The whole idea of microconsciousness is that processing in individual visual areas possesses a high degree of autonomy. Thus Zeki questions the increasingly popular view that projection to the fronto-parietal network is required for conscious experience (see, e.g. Rees et al 2000; Dehaene & Naccache 2001). Zeki has also provided striking evidence that processing of different visual attributes occurs on different time scales, from which he infers that we become conscious of visual attributes at different times (for an overview see, Bartels & Zeki 1998). We see the colour of an object 80ms before we see its motion, and we see its location before we see its colour (Moutoussis & Zeki 1997; Pisella et al 1998). On the basis of this evidence, Zeki claims that experiences of different sensory attributes are correlated with processing in specialised areas. These different experiences are each micro-consciousnesses that get bound together only after the processing required for experience has already taken place. For Zeki then, the neural representational systems that constitute the correlates of the visual experience of individual attributes are to be found in discrete, well-circumscribed areas of the brain.[10] Notice that an account of creature consciousness forms no part of Zeki's theory. He sets out an account of the neural basis of visual experience of particular visual attributes that

is quite independent of any explanation of creature consciousness or of the minimal sense of self.

A more tentative endorsement of localism can be found in Crick and Koch (2003). Crick and Koch claim that NCCs are most likely coalitions of pyramidal neurons located in the cerebral cortex. (Pyramidal neurons are a type of neuron that can communicate across large distances in the brain. They are capable of reaching from the back of the brain, for instance, where early visual processing takes place, to areas in the front of the brain where executive functions such as working memory and planning are carried out.) Each coalition will be made up of around a million neurons from the 50–100 billion neurons of which our brains are composed. Crick and Koch call the parts of a coalition "cortical nodes", each of which, in the context of the coalition of which it is a part, encodes an aspect of a percept. At any given time the brain will be made up of many such coalitions in competition with each other for a chance to influence further processing. When we attend to a given stimulus, a coalition sustains itself and suppresses competing coalitions. As our attention shifts to a new stimulus, so a different coalition will become dominant, suppressing activity in the earlier coalition, which immediately fades from awareness. A coalition realises a conscious experience when it reaches a certain threshold of activity for a certain time. The winning coalition is correlated with the contents of consciousness for as long as it suppresses the activity of competing coalitions. Koch & Tsuchiya (2005) for instance projected a faint, grey angry face to subject's right eye while flashing a stream of constantly changing colour patches to the other eye. The experience of the angry face image was completely suppressed even though it would have been clearly visible had the subject blinked their left eye. The coalitions of neurons responding to the angry face stimulus are suppressed by the dominant coalition responding to the stream of colour patches. Two different stimuli are presented to each eye, but which of the two stimuli ends up being seen will depend on which of the coalitions of neurons responding to the stimuli wins the competition and gets to sustain itself.

Crick and Koch hypothesise that the coalitions of neurons that can account for the different contents of consciousness are located in the cerebral cortex. They accept that these coalitions form parts of vast networks, and cannot perform their function apart from this context. Koch (in Koch & Greenfield 2007) for instance briefly discusses the case of Terri Schiavo who fell into a persistent vegetative state having undergone profound damage to the brain. When a person is in a deep coma, what are known as "arousal circuits" in the brain stem and thalamus are silent. Koch suggests that no stable coalitions can form in cerebral cortex without activity in these arousal circuits. On one natural reading, Koch can be taken to be claiming that activity in arousal circuits is a causal background condition rather than a constitutive background condition. Brainstem and thalamic activity, for instance, are correlated with conscious experience, on this picture, because activity in these areas is required for the formation of coalitions in the cerebral cortex. Brainstem and thalamic activity are required for the activation of the relevant cortical nodes, but don't form a part of the system that explains the contents of consciousness at a given time. On an alternative reading, activity in sub-cortical arousal circuits is a

10
Zeki distinguishes between micro-consciousness, macro-consciousness (consciousness of a bound visual percept) and unified consciousness. I am discussing the account he gives of the neural correlates of micro-consciousness.

constitutive enabling condition, and as such is a part of the minimal neural representational system that explains the contents of consciousness at a given time. Crick and Koch (2003) official position seems to be neutral on which of these two readings is correct (also see Rees et al (2002: 261)).

Consider as a final example of localism, Tong et al's (1998) binocular rivalry experiment in which subjects were presented with a picture of a face to one eye and a picture of a house to their other eye. Subjects reported that their conscious experience shifted every few seconds between a visual experience of the face image and an experience of the house image. Tong and colleagues used fMRI to record activity in the fusiform face area (FFA) and the parahippocampal place area (PPA). FFA responds twice as strongly to faces as to other stimuli, while PPA responds strongly to place-related stimuli. They found a strong correlation of activity in FFA when subjects reported their percept flipping to an experience of a face-image, and strong correlation with activity in PPA when subjects reported experiencing a house image. Moreover, a decrease in activity was observed in the respective areas when the preferred stimulus for that area popped out of awareness. This experiment seems to demonstrate that activity in one of these two cortical areas is necessary for either seeing a face-image or seeing a house-image. Activity in these areas constitutes the core realiser of such experiences. Nancy Kanwisher (2001), one of the experimenters in the original Tong binocular rivalry study, argues that PPA and FFA activity will however, not prove sufficient for experience. She claims that it is only in the context of processing that binds together representations to a particular time and place that activity of this kind can contribute to visual experience. According to Kanwisher, conscious experience may be the outcome of interactions "between domain-specific systems for representing the contents of awareness (primarily in the ventral visual pathway) and domain-general systems (primarily in the dorsal pathway) for organising those contents into structured percepts" (p. 109).

In what sense is Kanwisher's hypothesis a localist account? Kanwisher stresses the importance of interaction between ventrally and dorsally located areas. Her account of the NCC takes experience to be correlated with activity in cortical areas – the ventral and dorsal pathways. Kanwisher, like other localists, identifies the NCC for a given visual experience with an experience's core realiser, which Kanwisher takes to be activity in the ventral pathway. The constitutive background conditions will be the domain-general systems located in the dorsal pathway that account for object-representation. This domain-general system does not account for the difference between the presence of consciousness and its absence. Thus like other localist theories, Kanwisher's assumes we can treat the problem of explaining state consciousness as a separate problem from that of explaining creature consciousness.

Now that we have a few examples of localist theories before us, it is time we considered to what extent they could contribute to answering the questions motivating the gap intuition. All three of the localist theories we have sketched above share a common goal: they seek to identify the correlates of the contents of particular experiences. However, as Searle (2000) points out, it will only make sense to pursue an explanatory project of this kind in a creature that we know to already be conscious. It would make no sense to look for the neural correlate of a particular experience in a subject that was in the midst of an absent seizure, for instance, and so incapable of having conscious experiences. Since localist theories presuppose that a creature is conscious, we cannot expect to learn from such theories what it is for a creature to be

conscious. We can expect to learn at best what the differences are, at the neural level, between states that are conscious and states that are not, where the states in question belong to a creature that is already conscious.

This criticism relates to a point I made at the outset of this section. Localist theories, focussing as they do on the diverse contents of consciousness, neglect what is common to different experiences occurring in the consciousness of one and the same creature. There is however something the diverse contents of consciousness share in common: each of them is a state, event or process *for* one and the same conscious creature. The same conscious creature is acquainted with each and every one of its experiences under a first-person mode of presentation. The creature is acquainted with each conscious state as its own. What a creature is acquainted with is something that varies with differences in the contents of experience, but what remains the same is the first-person mode of givenness. Sometimes this acquaintance one has with one's own experiences will be very obvious as when one hears a loud, obtrusive sound or when one tastes a lemon. On other occasions this acquaintance will be something more subtle, and hence less attention grabbing. This acquaintance will however persist whether or not it becomes the object of one's focal attention. I suggested, in my analysis of creature consciousness, that a creature that has this kind of acquaintance with its mental states also has a minimal sense of self. It is this that is taken for granted by localist theories. I will argue next that not only do localist theories fail to account for this minimal sense of self. It is something they cannot account for without collapsing into holist theories.

We have just seen how localists fail to account for what is common to different experiences. Suppose the localist accepts the analysis I have given of the what-it-is-likeness of experience. She accepts that there is something an experience is like for a creature when that creature is acquainted under a first-person mode of presentation with the experience in question. Then she must take creature consciousness to be not just a causally necessary background condition, but what I have called, following Block (forthcoming), a "constitutive" background condition. Creature consciousness and its neural basis cannot just be making a causal contribution to experience, as localist theories must claim. Rather it forms a part of every experience explaining what it is for an experience to be an experience for a creature. The localist ought to accept that the neural activity underlying creature consciousness is a constitutive background condition. It is hard to understand however what distinguishes the parts of a total realiser that count as constitutive background conditions from the parts that count as an experience's core realiser. If some background conditions play a constitutive role in realising an experience, the conditions surely also count as parts of the total realiser for that experience. It is the fact that the core realiser plays a constitutive role in realising some state that is supposed to distinguish the core from the non-core parts of a total realiser. To claim that the neural activity that supports creature consciousness forms a part of the core realiser of a given experience is however just what the holist claims. Thus the localist faces a dilemma. She could reject the analysis that I've given of phenomenal consciousness, and deny that there is anything common to different conscious experiences, which explains what it is like for a creature to undergo these various experiences. Alternatively she could recognise that different experiences of the same creature have something in common, they each embody a minimal sense of self. However this is to give up on the idea that the neural basis of creature consciousness forms a caus-

ally necessary background condition. It is instead to recognise that the neural substrate underlying creature consciousness forms a core part of the structure of every conscious experience. It is of course the second of these options that I will pursue in the remainder of this paper.

4. The holist's account

Holists deny that the neural basis of creature consciousness is a causally enabling background condition. They claim that creature consciousness is a structural feature of every conscious experience. The neural basis of creature consciousness therefore forms a part of the core realiser of each and every conscious experience. Let us suppose that we can identify neural activity that forms the basis for creature consciousness. This will be neural activity that accounts for the minimal sense of self – for the contrast between consciousness and its absence. The holist claims that an account of the subject's being in particular conscious state will have to explain the integration of those neural representational systems that are correlated with creature consciousness and those correlated with particular states of consciousness. Before I offer an example of a holist theory, I will briefly explain how neuroscientists have gone about studying creature consciousness.

We have seen how neuroscientists study state consciousness by investigating under different experimental conditions what the neural differences are between states that are conscious and those that are not. The binocular rivalry studies are a nice example of this method. When the subject is conscious of the house-image she is not conscious of the face-image. By measuring brain activity during exposure to rivalrous imagery we can identify brain fluctuations specifically related to the contents of particular experiences. In a similar fashion, neuroscientists have studied creature consciousness by investigating the differences that occur in the brains of creatures that are conscious and creatures that are not.

One difficulty we immediately encounter is how to conceptualise the difference between the presence and absence of consciousness from a third-person or scientific perspective. We've already seen in §1 that it is not right to identify consciousness with wakefulness, since wakefulness and consciousness come apart in a number of cases, from patients in a vegetative state to patients suffering from epileptic seizures. I have suggested that we understand the contrast between the presence and absence of consciousness in terms of whether or not the creature has a minimal sense of self. In a sense however this is just renaming the problem. Each of us knows from the inside whether or not this minimal sense of self obtains, but this doesn't give us a third-person take on how to assess whether a creature, on a given occasion, has this sense of self. It doesn't tell us what we would need to do to test for the presence or absence of consciousness.

Damasio (1999) suggests a way of resolving this problem in his discussion of akinetic mutism (AM). Patients suffering from AM are unable to initiate voluntary movements, and cannot speak. They show some signs of attentive wakefulness, tracking the movement of their doctor as he move about their room for instance, but they exhibit no other signs of normal behaviour (Damasio 1999: 102; Schiff 2007: 592). These patients could move purposefully (they are not paralysed), but they lack the capacity to formulate plans on the basis of what they are sensing. They lack this ability to form intentions because they lack consciousness of what they are sensing. Further support

for the claim that there is a conceptual connection between the capacity to act intentionally and consciousness comes from absence automatism. Patients in the midst of an absence seizure exhibit motor behaviours of the type you might observe in a sleepwalker. They can walk about, take a drink from a glass, and in extreme cases may even leave the building and wander off down the street completely oblivious as to what they are doing. All of the behaviour we observe in such patients is carried out automatically without the formation of anything akin to an intention or plan. These patients cannot behave in anything other than an automatic fashion because they are unconscious. I suggest then that the capacity to form intentions and the minimal sense of self go hand in hand.[11] When we are acquainted with our experiences in a first-person way, it is this acquaintance that enables us to form intentions based on what we are experiencing. A sign of the absence of such a minimal sense of self is therefore to be found in subjects that cannot form intentions and plans based on what they perceive.

Now that we have a firmer idea of how to establish the absence of consciousness, let us return to the question of how neuroscience might go about finding its neural basis. One way to proceed would be to investigate what goes on in the brain during disturbances in consciousness (for a review see Schiff 2007). The classic work of Penfield and Jaspers (1954) with epileptic patients falls into this category. Penfield and Jaspers attempted to cure their patient's intractable epilepsy by removing large areas of cortex. They carried out these procedures under local anaesthesia, eliciting reports from the patients throughout the procedure. Amazingly, they found that the subject's continuity of consciousness remained undisturbed throughout the procedure. The removal of large areas of cortex deprived the patients of access to "certain forms of information, discriminative capacities, or abilities, but not of consciousness itself" (Merker, 2007: 5). This would seem to show that while cortical activity undoubtedly plays a central role in realising particular experiences, it might not be what accounts for the difference between the presence and absence of consciousness.

General anaesthetics are used to induce a general state of unconsciousness. One way to answer our question would therefore be to identify a common neural mechanism that explains how anaesthetics induce unconsciousness. Alkire et al (2000) used PET (positron emission tomography) to study the effects of two different anaesthetics on the brain. They found that both types of anaesthetic caused specific "reductions of regional cerebral glucose metabolism primarily in the thalamus and also in the midbrain reticular formation, basal forebrain, cerebellum, and occipital cortex" (p. 375). They hypothesise that anaesthetics "hyperpolarise" thalamocortical cells that normally transmit sensory information through the thalamus to cortical areas. (Cells are "hyperpolarised" when the chance of these cells transmitting a signal to other cells

11
Patients with locked-in syndrome may be thought to constitute a counterexample to this proposal. These patients are conscious unlike patients with akinetic mutism, but like the latter patients they cannot initiate voluntary movement with the important exception of blinking. Eye-movements in the vertical direction and blinking are the only remnants of the patient's ability to act. Are these patients, examples of consciousness in the absence of the ability to plan based on current perceptual information? It doesn't seem to me that they are. These patients can still form plans; what they have tragically lost is the ability to execute their plans. We only have to read Jean-Dominique Bauby's (1998) moving account of his experience of locked-in syndrome recounted entirely through blinking to know that these patients are certainly capable of forming intentions and plans.

is reduced.) This results in cortex becoming "functionally disconnected" from sensory information.

Alkire and his colleagues take their finding to provide support for the hypothesis that the neural substrate of consciousness involves thalamocortical-corticothalamic loops (see, e.g. Crick 1994; Llinas, Ribary, Contreras & Pedroarena 1998). According to this hypothesis, the thalamocortical system acts as a hub through which thalamic nuclei communicate with sensory, motor and associational cortical areas, and vice versa. A thalamo-cortical circuit that involves pyramidal neurons in layer IV of the neo-cortex, thalamic neurons, and reticular nucleus neurons is assigned the role of sustaining 40-Hz oscillations. This electrophysiological activity functions as the glue that binds together processing of attributes in different sensory areas. We have here our first example of a holist theory. According to this hypothesis, it is the thalamo-cortical loop as a whole that accounts for a given conscious experience. However only a part of the loop consists of activity relating to specific content – namely the sensory and motor neurons that project to layer IV of the cortex. The remainder of the loop (i.e. the intralaminar nuclei located in the thalamus) consists of non-specific activity that has the function of maintaining the cortex in an awake and alert state, making possible conscious experience, thought and deliberate action. It is both the specific and non-specific neural activity that accounts for a creature's being in a particular conscious state. On this proposal, an account cannot be given of the neural mechanisms that form the basis of a specific experience also act as the substrate of creature consciousness. I doubt, however, that the appeal that is made to the thalamo-cortical circuit is sufficient to explain what I have characterised as the minimal sense of self. Notice that the role that is assigned to the intralaminar nuclei is that of maintaining the cortex in an awake and alert state. However I've argued above that the minimal sense of self doesn't just consist in alertness and awakeness, since both can be present to some degree without a sense of self. Perhaps we will make better progress by considering what occurs in the brain in the other examples of disturbances of consciousness I gave above.

Let us turn our attention to some longer lasting, often more fatal disturbances of consciousness, namely coma and the persistent vegetative state. Both of these disturbances in consciousness are the result of damage to a small region of the brain stem, the part of the brain that connects the spinal cord to the cerebral hemispheres. The brain stem is composed of numerous small nuclei (or three-dimensional collections of neurons), and interconnecting nerve fibres. Each nucleus has its own distinctive cellular composition ("cytoarchitecture") and neurochemical identity. Moreover, the nuclei that make up the brain stem have diverse functions and project to distinct sets of neural structures. The core region of the brain stem, the so-called *reticular formation* projects to the intralaminar nuclei in the thalamus, which as we have just seen, may play a necessary role in realising creature consciousness. The damage to the brain stem that causes coma and vegetative state encompasses nuclei found within the reticular formation.

It was traditionally thought that the function of the reticular formation was to wake up and energise the thalamus and cerebral cortex. Such a view coheres well with the account of the thalamo-cortical circuit briefly sketched above. This view has however been called into question by increasing evidence of the heterogeneity of the nuclei to be found in the reticular formation (for an extensive review of the literature see Parvizi & Damasio 1999). Each nucleus may have a different role to play in modulating activity in the cerebral cortex.

On the basis of this evidence Damasio (1999: ch.8) has argued for a different, but complementary, understanding of the functional role of the reticular formation. Damasio accepts that activity in the reticular formation may cause the coherent, local and global electrophysiological activity in the cerebral cortex characteristic of the wakeful and attentive state. However, he proposes that in addition many of the nuclei to be found in the reticular formation receive signals from the body that carry information about the state of the organism at the time. This part of the brain stem serves as an entry point for neural and chemical signals travelling to the brain from the central nervous system. Nuclei in the reticular formation project to other areas of the brain such as the hypothalamus and the cerebral cortex that use these signals to construct a detailed map of the bodily states of the organism from moment to moment. This map can then be used to regulate bodily functions required to keep the organism alive. Damasio calls this map of the organism's bodily states, the *proto-self*.

When damage occurs to the reticular formation of the kind that result in coma and vegetative state the brain can no longer monitor the organism's internal bodily states. Damasio writes:

> "I see one powerful fact emerging about the critical region of the brain stem we have been discussing: it is simultaneously engaged in processes concerning wakefulness, homeostatic regulation, emotion and feeling, attention and consciousness… Homeostatic regulation, which includes emotion, requires periods of wakefulness (for energy gathering); periods of sleep (presumably for restoration of depleted chemicals necessary for neuronal activity); attention (for proper interaction with the environment); and consciousness (so that a high-level of planning of responses concerned with the individual organism can eventually take place)." (1999: 260)

An organism needs to be in constant receipt of information about the state of its body if it is to enjoy anything in the way of consciousness. Deprived of this information by damage to the part of the brain that relays this information, and the result will be that the organism is also deprived of consciousness. However, Damasio argues that normal reticular activity isn't sufficient for the presence of consciousness. In addition the brain must map the relations between the organism and object. Any given representation of an object will cause changes in the organism's bodily states, which in turn will be mapped by the brain. Damasio suggests that the way for the brain to represent the relation of the organism to an object would be for the brain to produce what he calls a "second-order map". These second-order maps function as what Damasio calls "nonverbal narratives", telling the story of how the organism's bodily states are modified by the objects it is representing. Damasio identifies the neural basis of core consciousness with activity in the cingulate cortex, the superior colliculi, and the thalamus. It is these areas that are responsible for mapping the organism's changing relation to objects.

Damasio's account is an excellent example of a holist theory. What we have been calling "creature consciousness" he calls "core consciousness". His account of core consciousness is therefore an explanation of the contrast between consciousness and its absence. However, the neural mechanisms that account for the specific contents of experience also feed into this account. Damasio characterises these mechanisms as having the function of supplying something-to-be-known to the organism. Damage to these mechanisms will have an impact on what the organism can represent. It may result in perceptual disorders that prevent a sensory representation in a particular modality from being formed. Alternatively, it may result in agnosia – an object of perception may be deprived of its meaning. Crucially however, the neural

mechanisms that account for what is represented only get to make a contribution to consciousness by being included in a second-order map – a higher-order representation of how the organism's internal bodily states are affected by the objects it perceives. It is this second-order map that will constitute the contents of consciousness from moment to moment. The neural mechanisms that account for the specific contents of experience are different from those that form the basis for core consciousness. However it is only by making connections with areas involved in the construction of the second-order maps (namely, the cingulate cortex, the superior colliculi, and the thalamus) that these neural mechanisms get to make a contribution to consciousness. It is the neural mechanisms that account for core-consciousness that get to determine the contribution a given representation of object makes to consciousness.

What I have earlier called the minimal sense of self has a place at the heart of Damasio's account. Might Damasio have identified the neural basis of pre-reflective self-consciousness, the sense of self a creature has when it is acquainted with its mental state under a first-person mode of presentation? Unfortunately, I think we must look elsewhere. Damasio explains the minimal sense of self by appealing to the contents of a certain type of neural representation, a second-order map. A creature has a minimal sense of self because its brain has produced a certain type of higher-order representation. However, on the account I have sketched above the minimal sense of self isn't something that is produced by the tokening of a higher-order representation, the result of a relation between two non-conscious neural representations. Rather, a minimal sense of self is a structural feature intrinsic to every experience. It's hard to understand how the brain's producing a second-order map of the kind Damasio appeals to, could account for the awareness I have of an experience as mine. It's hard to understand how introducing an additional layer of neural representation could transform a representation into one a creature is acquainted with under a first-person mode of presentation. In the next and final section I will argue that a dynamic sensorimotor account of consciousness may well fare better.

5. Towards a DSM account of the minimal sense of self

Whereas localists view neural processing taking place in particular sensory systems as happening independently of motor and autonomic processing, this is an assumption that is rejected by holist theories. Holists stress the integration of different types of sensory information. Damasio writes:

> "There is no such thing as a *pure* perception of an object within a sensory channel, for instance, vision. The concurrent changes I have just described are *not* an optional accompaniment. To perceive an object, visually or otherwise, the organism requires both specialised sensory signals *and* signals from the adjustments of the body, which are necessary for perception to occur." (1999: 147)

On Damasio's account, creature consciousness is the result of the integration of the information from these multiple sensory channels in a second-order map. We can explain the specific contents of consciousness from moment to moment only by identifying the neural mechanisms that account for the integration of these different types of sensory information. In this section I will go a step further and suggest that the specific content and character of experience may be the result of a certain kind of integration, namely sensorimotor integration. I will suggest that the neural mechanisms that underlie sensorimotor

integration may also form the basis for a minimal sense of self. Since sensorimotor integration explains the contents of experience and may also explain the sense of self, the result is a holist theory: a neural description of state consciousness is given as a part of a theory of creature consciousness.

The DSM account of consciousness claims that perception isn't just the passive reception of information, but is rather an active process of gathering and assembling information from the environment. Perceiving is a skillful bodily activity in which a certain kind of know-how is exercised. Just as riding a bicycle is a skill which one possess when one has a certain kind of know-how (i.e. one knows-how to ride a bike), so perceiving is a skill which one can exercise when one has the relevant know-how. In the case of perception one must be able to track an object or property across changes in sensory input brought about through movement of the relevant sense organ, head, upper torso and whole body. A subject that is able to track an object or property across such changes has what I shall call *sensorimotor knowledge.*

As mobile animals, we have access to (in the case of vision) dynamic flows of continuously varying retinal information. We have information about how such optic flows vary as a function of movement (see O'Regan and Noë 2001; Noë 2004). We can discern in these information flows, invariant structures corresponding to objective properties of things in our surrounding environment. As we move around a rectangular table, for instance, the information our eyes receive from the table varies but it doesn't vary in a chaotic or random fashion. Rather, there is structure to the way the flow of information varies, the structure characteristic of a rectangular object. There is a relation of lawful dependence that holds between movement and the visual profiles the rectangular object presents. A perceiver that has an implicit understanding of laws relating information flows to movement will be able to perceive the table's rectangularity.

O'Regan and Noë call the relations of lawful dependence that hold between sensory flows of information and movement, *patterns of sensorimotor contingency.* The idea is that there are patterns or invariant structures in sensory experience contingent on movement. The DSM account claims that all perceptual experience is mediated by implicit knowledge or understanding of these patterns of sensorimotor contingency. Its central thesis is that perceptual experience just is the exercise of sensorimotor knowledge. The core idea as I understand it is that perceivers that can exercise sensorimotor knowledge can pick up on invariant structures in sensory flows of information. Since perception consists in immediate and direct pick-up of such information, it is the possession and exercise of sensorimotor knowledge that accounts for the perception of objects and their sensible properties. It is the possession and exercise of sensorimotor knowledge that explains how we can immediately and directly pick-up information about objects and their sensible properties.

The DSM account purports to offer an account of the content and character of experience in terms of patterns of sensorimotor contingency. Our experiences of, for instance, red things have a particular reddish quality to them because of the particular sensorimotor contingencies or laws of sensorimotor dependence that govern such experiences. These laws determine amongst other things how the surface of an object will cause different sensory stimulation under different illumination conditions and the effects of our movements on the sensory inputs red objects typically cause in us (for much more on the case of colour see Noë 2004: ch. 4). Other sensorimotor contingencies purport to explain the character of our experience rather than its content; the difference

between sense modalities, e.g. the difference between seeing and touching. As you move your head, for instance, the objects one is seeing come into and go out of view. Forward movement brings about an expansion in optic flow while backward movement generates contraction. Blinking, turning away, and shutting of the lights has the effect of terminating one's visual contact with an object. Touch, smell and hearing will not be affected in these ways by movement but will be governed by a different set of sensorimotor contingencies. As you move your hand away from an object, nothing analogous with a contraction of the visual field happens. Information is picked up using touch in a very different way from vision. It is true that at a certain level of abstraction we can map the sensorimotor contingencies governing touch onto those governing vision and vice versa (see Noë's 2004, §3.8 discussion of Molyneux's question). Noë writes:

> "If something looks square, then one would have to move one's eyes or head in characteristic ways to look at each of its corners. One would have to move one's hands in the same way (at the appropriate level of abstraction) to feel each corner." (2004: 102)

Nevertheless it remains the case that the ways looks vary with movement is very different from the ways tactile sensations vary. This, says Noë, gives us an account of the difference in character between seeing and touching.

Thus the DSM account seems to have a promising story to tell about why our experiences have the content and character they do. I want to suggest next that it also contains the seeds of an account of the minimal sense of self. The minimal sense of self is the result of the integration of sensory flows of information with re-afferent information – information relating to one's own movements. The DSM account claims that there will typically be three types of information involved in sense experience:

(1) Information that relates to motor plans or intentions to move.
(2) Predictions about the motor and sensory consequences of carrying out a particular motor plan. Such predictions will be fuelled by what I have earlier called "sensorimotor knowledge".
(3) Actual sensory input, including proprioceptive and kinaesthetic feedback.

According to the DSM account, the neural basis of any given experience will consist of the neural activity involved in processing these three types of information. To account for the minimal sense of self we need to suppose that there is a mechanism in the brain that integrates these different types of information. When this mechanism finds a coherence between information of types (2) and (3) – the predictions that are made on the basis of sensorimotor knowledge about the sensory consequences of movement and the actual sensory input our perceptual systems receive – the resulting experience will feel like our own experience. It will be *given* as our own. It is precisely this sort of givenness that is the defining characteristic of the minimal sense of self. Our experiences incorporate a minimal sense of self because they are given as our own. I suggest that it is the coherence between these three types of information, which forms the basis for a minimal sense of self. On the DSM account, experiences have a particular content that is the result of the integration of these three different types of information. Thus the DSM account can readily explain how the minimal sense of self can be built into every experience. Moreover, the DSM account as I have sketched it will qualify as a holist theory. The account it gives of the contents of particular experience is inseparable

from the account it gives of creature consciousness, or what I have called the minimal sense of self.

Proponents of a DSM account have hitherto steadfastly opposed the NCC program. They have combined a defence of the DSM account with a commitment to vehicle externalism (see, e.g. Hurley 1998; Noë 2004; 2006; 2007). Vehicle externalists claim that as embodied subjects we are tightly coupled to the world so that we cannot simply "unplug" the brain from the body and the environment. The body and the environment are affecting and constraining brain processing to such a great extent that it is simply not possible for the very same brain processing to take place in a different bodily or environmental setting. Noë makes the following analogy:

> "The states of a car's engine are necessary conditions of its driving activity; moreover, in certain conditions one can change the car's driving behaviour by directly modulating the states of its engine. But it is absurd to think that the states of the engine are alone sufficient for driving! The engine needs to be properly embodied in the vehicle, and the car itself must be situated in an appropriate environment. A car suspended from a hook, or up to its windows in mud, won't drive, no matter what the state of the engine." (2004: 211)

One might suppose that vehicle externalism of this kind is incompatible with what I have earlier called "the minimal supervenience thesis". This thesis says that for any given experience *e* there exists a neural system – a configuration of neural activity – that constitutes the minimal supervenience base for the occurrence of *e*. We might take vehicle externalists to be saying that the minimal supervenience base for some of our experiences extends outside the brain. The holist account I have proposed is however quite consistent with vehicle externalism. It can happily accept that body and environment might be continuously affecting and constraining brain processing. It takes the NCC for a given experience to consist of the neural activity underlying the three types of information described above, and the mechanisms responsible for integrating these types of information. It is not however committed to the possibility of the same brain processing taking place in a different bodily or environmental setting. A holist theory of the kind I have sketched could allow that there is no unplugging the brain from its bodily and environmental setting.

Noë and Thompson (2004) deny that neural representational systems could possibly have a content that matches that of conscious experience. They accept the possibility that the contents of neural representational systems may *agree* with the contents of experience but they deny that neuroscientists could ever find an exact match. Thompson (2007) writes:

> "Experiential content and neural content are different kinds of content, and so it is a category mistake to confuse the two. Experience is intentional (world-presenting), holistic (constituted by interrelated perceptions, intentions, emotions and actions), and intransitively self-aware (has a nonreflective subjective character)." (p. 350)

The holist account of NCCs I have been developing can however accommodate all three of the features Thompson rightly takes to be definitive of experiential content. It can accept that experience is *world-presenting*. We have seen how the DSM account explains the world-presenting contents of experience in terms of patterns of sensorimotor contingency. A holist theory along the lines described above explains how such patterns of sensorimotor contingency could be embodied in neural activity. My account can also accommodate the fact that experience is *holistic.* The DSM account claims that the contents of any given sensory experience will include information relating to motor intentions, predictions about the sensory consequences of carrying

out such a motor intention and actual sensory input, including presumably feedback relating to how the organism's internal bodily states are affected. Finally what Thompson calls *intransitive self-awareness* is what I have been calling a minimal sense of self. We have also seen how the DSM account can explain the minimal sense of self. Thus a holist account of NCCs along the lines I have sketched above would seem to provide an account of neural content that has all of the features characteristic of experiential content.

Do the body and environment play a *constitutive* role in realising experience or do they only make a *causal* contribution? One way of bring out what is it as issue here is to ask if a duplicate of my brain were to suddenly come into existence free-floating in space would this duplicate brain instantiate any of my experiences? We might say that it doesn't matter what sort of environment the brain is hooked up to, if this brain duplicates the structure and organisation of my brain it will have experiences just like mine at the moment it comes into existence. Body and environment make a merely causal contribution to brain processes. We can rig things up so that the same contribution can be made by something other than the body and environment. We have already seen however that a DSM account along the lines I have sketched has little reason to accept this as anything other than a metaphysical possibility. If the duplicate brain isn't coupled with an environment like our own, there is nothing to guarantee it will have experiences that are anything like our own. Hurley (forthcoming) makes a very similar point in her discussion of supervenience thought experiments (STEs). STEs invite us to hold constant internal factors such as neural activity whilst varying external factors such as the environment of the agent. They invite us to "unplug" the internal factors from one environmental setting and replug the very same internal factors into a different environmental setting. STEs then proceed to ask whether the content or quality of experience will likewise vary. This presupposes that the internal factors can be unplugged from the external factors, something the vehicle externalist denies. I see it as an empirical question to be decided on a case by case basis the extent to which this sort of unpluggability holds.

On the localist conception of NCCs, body and environment can at best make a causal contribution to realising experience. This is not so on the holist account of NCCs I have been developing. On this account both creature consciousness and the contents of consciousness are the result of dynamic relations between sensory flows of information, movement and feedback from movement. These different flows of information will form the content of a single experience as the result of the integration and coordination of neural activity in diverse areas of the brain. There is however no reason to suppose that the brain processing on its own without the relevant bodily and environmental setting could realise experience as we know it.

References

Alkire, M. T., Haier, R. J. & Fallon, J. H. 2000: "Towards a unified theory of narcosis: Brain-imaging evidence for a thalamocortical switch as the neurophysiologic basis of anaesthetic-induced unconsciousness". In *Consciousness and Cognition,* 9: 370–386.

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

Bartels, A. & Zeki, S. 1998: "The theory of multi-stage integration in the visual brain". In *Proceedings of the Royal Society of London Series B,* 265: 2327–32.

Bauby, J-D. 1998: *The Diving Bell and the Butterfly: a memoir of life in death* (Vintage).

Bayne, T. 2007: "Conscious states and conscious creatures: Explanations in the scientific study of consciousness". In *Philosphical Perspectives* 21: *Philosophy of Mind*, 1–22.

Block, N. forthcoming: "Consciousness, accessibility, and the mesh between psychology and neuroscience". In *Behavioural and Brain Sciences.*

Block, N. 2005: "Two neural correlates of consciousness". In *Trends in Cognitive Science,* 9.2: 46–52.

Block, N. & Stalnaker, R. 1999: "Conceptual analysis, dualism, and the explanatory gap". In *Philosophical Review,* 108: 1–46.

Chalmers, D. 2002a: "Does conceivability entail possibility?". In T. S. Gendler & J. Hawthorne (Ed.'s) *Conceivability and Possibility,* Oxford: Clarendon Press.

Chalmers, D. 2002b: "Consciousness and its place in nature". In D. Chalmers (Ed.) *Philosophy of Mind: Classical and Contemporary Readings,* Oxford: Oxford University Press.

Chalmers, D. 2000: "What is a neural correlate of consciousness?". In T. Metzinger (Ed.) *Neural Correlates of Consciousness: Empirical and Conceptual Issues,* Cambridge, MA: MIT Press.

Chalmers, D. 1996: *The Conscious Mind,* Oxford: Oxford University Press.

Churchland, P. S. 2005: "A neurophilosophical slant on consciousness research". In *Progress in Brain Research,* 149: 285–93.

Cosmelli, D. & Thompson, E. "Embodiment or envatment: reflections on the bodily basis of consciousness". In J. Stewart, O. Gapenne, and E. di Paolo (Ed.'s) *Enaction: Towards a New Paradigm for Cognitive Science,* Cambridge: MIT Press.

Crick, F. 1994: *The Astonishing Hypothesis: the Scientific Search for the Soul,* New York: Scribner.

Crick, F. & Koch, C. 2003: "A framework for consciousness". In *Nature Neuroscience,* 6.2: 119–126.

Damasio, A. 1999: *The Feeling of What Happens: Body, Emotion and the Making of Consciousness,* London: William Heinemann.

Dehaene, S. & Naccache, L. 2001: "Towards a neuroscience of consciousness". In *Cognition,* 79: 1–37.

Dennett, D. 1991: *Consciousness Explained,* New York: Little Brown.

Dretske, D. 1995: *Naturalising the Mind,* Cambridge, MA: MIT Press.

Harman, G. 1990: "The intrinsic quality of experience". In *Philosophical Perspectives,* 4: 31–52.

Hilll, C. & McLaughlin, B. 1999: "There are fewer things in reality than are dreamt of in Chalmers' philosophy". In *Philosophy and Phenomenological Research,* 59: 445–54.

Hohwy, J. (In Progress) "A unified framework for the neural correlates of consciousness".

Hohwy, J. 2007: "The search for neural correlates of consciousness". In *Philosophy Compass,* 2/3: 461–74.

Hurley, S. forthcoming: "Varieties of externalism". In R. Menary (Ed.) *The Extended Mind* (Ashgate Press).

Hurley, S. 1998: *Consciousness in Action,* Harvard: Harvard University Press.

Kanwisher, N. 2001: "Neural events and perceptual awareness". In *Cognition,* 79: 89–113.

Koch, C. & Greenfield, S. 2007: "How does consciousness happen?". In *Scientific American.*

Lamme, V. A. 2004: "Separate neural definitions of visual consciousness and visual attention: a case for phenomenal awareness". In *Neural Networks,* 17: 861–72.

Laureys, S. 2005: "The neural correlates of (un)awareness: lessons from the vegetative state". In *Trends in Cognitive Science,* 9: 556–559.

Legrand, D. 2007: "Pre-reflective self-as-subject from experiential and empirical perspectives". In *Consciousness and Cognition,* 16: 583–99.

Llinas, R., Ribary, U., Contreras, D., & Pedroarena, C. 1998: "The neuronal basis of consciousness". In *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 353: 1841–49.

Loar, B. 1997: "Phenomenal States". In N. Block, O. Flanagan, and G. Güzeldere (Ed.'s) *The Nature of Consciousness: Philosophical Debates,* Cambridge, MA: MIT Press.

Mackie, J. L. 1974: *The Cement of the Universe,* Oxford: Oxford University Press.

McLaughlin, B. 2007: "Type Materialism for Phenomenal Consciousness". In M. Velmans, S. Schneider (Ed.'s) *The Blackwell Companion to Consciousness,* Oxford: Blackwell.

Merker, B. 2007: "Consciousness without a cerebral cortex: a challenge for neuroscience and medicine". In *Behavioural and Brain Sciences,* 30: 63–81.

Moutoussis, K. & Zeki, S. 1997: "Functional segregation and temporal hierarchy of the visual perceptive systems". In *Proceedings of the Royal Society of London B: Biological Sciences,* 264: 1407–14.

Noë, A. 2007: "Magic Realism and the Limits of Intelligibility: What makes us Conscious". In *Philosphical Perspectives,* 21: *Philosophy of Mind*: 457–74.

Noë, A. 2006: "Experience without the head". In T. Gendler & J. Hawthorne (Ed.'s) *Perceptual Experience,* Oxford: Oxford University Press.

Noë, A. 2004: *Action in Perception,* Cambridge, MA: MIT Press.

Noë, A. & Thompson, E. 2004: "Are there neural correlates of consciousness?". In *Journal of Consciousness Studies,* 11.1: 3–28.

O'Regan, J.K. & Noë, A. 2001: "A sensorimotor account of vision and visual consciousness". In *Behavioural and Brain Sciences,* 24: 939–1011.

Papineau, D. 2002: *Thinking about Consciousness,* Oxford: Oxford University Press.

Parvizi, J. & Damasio, A. 2001: "Consciousness and the brainstem". In *Cognition,* 79: 135–60.

Pascual-Leone, A. & Walsh, V. 2001: "Fast backprojections from the motion to the visual area necessary for visual awareness". In *Science,* 292: 510–512.

Penfield, W. & Jaspers, H.H. 1954: *Epilepsy and the functional anatomy of the human brain,* London: J. & A. Churchill Ltd.

Perry, J. 2001: *Knowledge, Possibility and Consciousness,* Cambridge, MA: MIT Press.

Pisella, L., Arzi, M., and Rossetti, Y. 1998: "The timing of colour and location processing in the motor context". In *Experimental Brain Research,* 121: 270–6.

Rees, G., Krieman, G., Koch, C. 2002: "Neural correlates of consciousness in humans". In *Nat. Rev. Neuroscience,* 3.4: 261–70.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., Driver, J. 2000: "Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction". In *Brain,* 123: 1624–33.

Rey, G. 1995: "Towards a projectivist account of conscious experience". In T. Metzinger (Ed.) *Conscious Experience* (Ferdinand Schöningh).

Rosenthal, D. 1993/2005: "Thinking that one thinks", reprinted as ch. 2 in *Consciousness in Mind,* Oxford: Oxford University Press.

Rosenthal, D. 1990/1997: "A theory of consciousness". In N. Block, O. Flanagan, and G. Güzeldere (Ed.'s) *The Nature of Consciousness: Philosophical Debates,* Cambridge, MA: MIT Press.

Ryle, G. 1949: *The Concept of the Mind,* London: Hutchinson & Co.

Schiff, N. 2007: "Global disorders of consciousness". In M. Velmans, S. Schneider (Eds) *The Blackwell Companion to Consciousness,* Oxford: Blackwell.

Searle, J. 2000: "Consciousness". In *Annual Review of Neuroscience,* 23: 557–578.

Shoemaker, S. 1981: "Some varieties of functionalism". In *Philosophical Topics,* 12: 93–119.

Sturgeon, S. 2001: *Matters of Mind,* London: Routledge.

Thompson, E. 2007: *Mind in Life: Biology, Phenomenology, and the Sciences of Mind,* Cambridge, MA: Harvard University Press.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. 1998: "Binocular rivalry and visual awareness in human extrastriate cortex". In *Neuron,* 21: 753–59.

Tsuchiya, N & Koch, C. 2005: "Continuous flash suppression reduces negative after-images". In *Nature Neuroscience,* 8: 1096–1101.

Tye, M. 2000: *Colour, Consciousness and Content,* Cambridge, MA: MIT Press.

Zahavi, D. 2006: *Subjectivity and Selfhood: Investigating the First-Person Perspective,* Cambridge, MA: MIT Press.

Zeki, S. 2007: "A theory of micro-consciousness". In M. Velmans, S. Schneider (Eds) *The Blackwell Companion to Consciousness,* Oxford: Blackwell).

Zeki, S. 2003: "The disunity of consciousness". In *Trends in Cognitive Science,* 7.5: 214–18.

Julian Kiverstein

Bewusstsein, minimales Selbst und Gehirn

Zusammenfassung

Dieser Artikel macht von der Möglichkeit Gebrauch, das Phänomen des Bewusstseins neurowissenschaftlich zu erklären, und geht der Frage nach, wie eine solche Erklärung wohl auszusehen hätte. Der Verfasser widmet sich konkret der These, dass jeder Erfahrung ein repräsentatives neurales System zugrunde liegt, das als Supervenienzgrundlage dieser Erfahrung dient. Diese Hypothese wird im weiteren Verlauf als minimale Supervenienz-These bezeichnet. Nach Meinung des Autors kann diese These auf zweierlei Weisen verstanden werden; dementsprechend ist von einer lokalistischen und einer holistischen Lesart die Rede. Lokalistische Theorien versuchen, die minimale Supervenienzgrundlage einer spezifischen Erfahrung zu ermitteln, während sie Fragen über die Natur des Bewusstseins von Lebewesen beiseite lassen und die Neuralbasis des Bewusstseins lediglich als kausal-notwendige Hintergrundvoraussetzung für eine partikuläre Bewusstseinserfahrung werten. Holisten hingegen räumen dem Bewusstsein von Lebewesen Vorrang ein und behaupten, dass partikuläre Bewusstseinszustände nur im Rahmen von Deutungen ebendieses Bewusstseins erklärt werden können. Nach Meinung des Autors muss jegliche wissenschaftliche Deutung des Bewusstseins jenem Faktor Rechnung tragen, den er als minimale Wahrnehmung des Selbst bezeichnet und das jedem Bewusstseinszustand intrinsisch ist. Holistische Theorien sind am besten geeignet, diesen Umstand zu verändern. Der Artikel schließt mit der These, dass die sog. Dynamisch-Sensomotorische (DSM) Deutung des Bewusstseins mit der holistischen Erklärung der Neuralbasis des Bewusstseins kombiniert werden kann. Eine solchermaßen kombinierte Sichtweise korrigiert den Widerspruch bezüglich der These von der minimalen Supervenienz, die bei einigen prominenten Befürwortern der DSM-These (z.B. Alva Noë und Evan Thompson) zu finden ist. Sie ermöglicht ebenfalls einen Rahmen für die Entwicklung einer neurowissenschaftlichen Deutung der These von der minimalen Wahrnehmung des Selbst.

Schlüsselbegriffe

Neurale Voraussetzungen des Bewusstseins, Zustandsbewusstsein, Bewusstsein von Lebenswesen, minimale Wahrnehmung des eigenen Selbst, vorreflektives Selbst-Gewahrsein, sensomotorische Dynamik

Julian Kiverstein

La Conscience, le Soi Minimal et le Cerveau

Resumé

L'article cherche à savoir si une explication neuroscientifique de la conscience est possible et à quoi elle pourrait ressembler. Plus particulièrement, je me pencherai sur l'affirmation qu'à chaque expérience donnée correspond un système de représentation neural qui constitue la base de survenance minimale de cette expérience. J'appellerai cette hypothèse « la thèse de survenance (supervenience) minimale ». Je soutiens que cette thèse peut se lire de deux façons que je nommerai lectures localiste et holiste. Les théories localistes cherchent à définir quelle est la base de survenance minimale des expériences particulières. Elles laissent de côté les questions sur la nature de la conscience des êtres et considèrent la base neurale de leur conscience comme une condition causale circonstancielle nécessaire à l'expérience d'une conscience particulière. Les holistes, d'autre part, donnent la priorité à la conscience des êtres et affirment que nous pouvons rendre compte des états de conscience particuliers seulement dans un contexte d'explication de la conscience des êtres. J'affirme que toute explication scientifique de la conscience doit rendre compte de ce que j'appellerai « une sensation minimale de soi », intrinsèque à tout état conscient. Les théories holistes sont les plus aptes à tenir compte de cet aspect. Enfin, j'affirme que l'explication sensori-motrice dynamique de la conscience peut se combiner avec une approche holiste de la base neurale de la conscience. Une telle combinaison d'approches compense l'opposition à la thèse de survenance minimale de certains défenseurs de l'explication sensori-motrice dynamique distingués (ex. Alva Noë et Evan Thompson). Elle offre également un cadre de développement pour une explication neuroscientifique de la sensation minimale de soi.

Mots-clés

corrélats neuraux de la conscience, état mental conscient, conscience de créature, sens minimal de soi, conscience pré-réflexive de soi, dynamique sensori-motrice