

# Functions of Consciousness<sup>1</sup>

Anil Seth

Department of Informatics, University of Sussex, Brighton, BN1 9QJ, UK

Email: [a.k.seth@sussex.ac.uk](mailto:a.k.seth@sussex.ac.uk)

Web: [www.anilseth.com](http://www.anilseth.com)

Tel: +44 1273 678549

Fax: +44 1273 678773

December 6, 2008

*“The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be completely without any power of modifying that working, as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery.” (Thomas Huxley, quoted in James, 1890, p.135).*

*“The particulars of the distribution of consciousness, so far as we know them, point to its being efficacious ... it seems an organ, superadded to other organs which maintain the animal in the struggle for existence; and the presumption of course is that it helps him in some way in the struggle ...” (James, 1890).*

## 1.0 Introduction

A major challenge for the successful naturalization of consciousness lies in locating its biological function, or functions. Although common sense suggests that conscious experience has many important functional roles in our lives, experiments and theoretical arguments challenge these everyday intuitions. Many human behaviors can occur in the absence of consciousness, and the natural world contains many creatures capable of engaging in complex behaviors, at least some of which may be doing so entirely without consciousness (e.g., mollusks, microorganisms). While consciousness is a real phenomenon whether functional or not, without any defensible function its scientific study is made even more difficult than it already is (Chalmers 1996; Humphrey 2002).

We begin with some necessary conceptual ground-clearing, reviewing the possibility that consciousness does not have a function and dealing with ambiguities in the meaning of ‘function’ and of ‘consciousness’. We then consider in detail a range of putative functions for consciousness, covering both cognitive perspectives and proposals grounded in neural dynamics. We start with the intuitively appealing ideas that consciousness functions to initiate voluntary behavior and/or to mediate rational actions, before turning to more recent candidates, for example that consciousness functions to

---

<sup>1</sup> This article is a fully referenced and extended version of a chapter of the same title, which will appear in the forthcoming *Elsevier Encyclopedia of Consciousness*, ed. W.P. Banks, Elsevier Press.

integrate information or to mediate planning and flexible behavior in response to novelty. Finally, we consider the possibility that consciousness, as a constellation concept (Block 1995; Block 2005; Zeman 2005) may have multiple functions.

## 2.0 Does consciousness have a function?

Asking about the function of consciousness requires first considering the possibility that consciousness has no function. There are two sorts of arguments along these lines; conscious inessentialism and epiphenomenalism.

### 2.1 Conscious inessentialism (CI)

This is the view “that for any intelligent activity  $i$ , performed in cognitive domain  $d$ , even if  $we$  do  $i$  with conscious accompaniments,  $i$  can in principle be done without these conscious accompaniments” (Flanagan 1992). CI is a radical proposition and one which certainly challenges common sense. However, since many biological functions can be carried out by a variety of mechanisms (Edelman & Gally 2001; Leonardo 2005), CI is not *obviously* false.

Some behaviors, for example introspecting, seem obviously to require consciousness. If introspection is defined as explicit verbal report of the content of conscious experiences, then introspection requires consciousness. However, examples like these are not strong challenges to CI because they represent cases in which consciousness is constitutively essential rather than causally essential (Dretske 1997). To the extent that introspection is defined in terms of consciousness the notion of unconscious introspection is simply incoherent.<sup>2</sup>

If CI is true then ‘zombies’ are possible; in fact, the possibility of zombies is just the thesis of CI embodied (Polger & Flanagan 1995). Zombies in philosophy of mind are creatures that are behaviorally indistinguishable from normal (conscious) humans but who are entirely without consciousness; there is no “what it is like” to be a zombie. While the zombie concept has served to emphasize the question of the function of consciousness, logic-based attempts to settle the conceivability or otherwise of zombies have remained controversial [see, for example (Chalmers 1996; Dennett 1991; Moody 1994)]. It has been argued that if zombies are conceivable in worlds that share our physical laws, then we can conclude that even in our world, consciousness does not causally influence the physical events responsible for human behavior (van Gulick 2004). But this inference is false. From the statement that consciousness is inessential for a particular behavior it does not follow that consciousness is causally ineffectual for that behavior in those cases when it is present (see Section 2.2).

---

<sup>2</sup> There are surprisingly few examples that fall unambiguously into this category. For example, *lying* might be considered constitutively essential if defined in terms of a conscious intention to mislead, as opposed to merely having the effect of misleading. However, it seems possible that lying could also be defined as the intention to mislead where the intention could be either conscious or unconscious.

A more stringent form of the zombie argument requires that zombies not only display identical behaviors to normal humans but that they also have complete neurophysiological equivalence (Chalmers 1996; Kirk 1974). However, such ‘brainy zombies’ may be conceivable only if one allows non-naturalistic explanations of consciousness. By all naturalistic accounts it seems impossible for there to exist two brains in exactly the same physical state but only one of which is conscious.

## **2.2 Epiphenomenalism (EP)**

This is the view expressed by Thomas Huxley in the introductory quote. According to the epiphenomenalist suspicion (EP), consciousness exists but does not play any causal role in neural or cognitive operations [see (Chalmers 1996; Robinson 2007) for a discussion].

Even if CI is false, this does not by itself imply that EP is also false. It is conceivable that a behavior/activity requires the sort of brain activity that inevitably gives rise to consciousness, without consciousness itself playing a causal role in generating that behavior. Conversely, even if CI is true it does not necessarily follow that EP is also true. As noted above, even if consciousness is inessential for a particular behavior it does not follow that consciousness has no causal role in generating the behavior in those instances in which consciousness is present. While according to James, EP is an “unwarrantable impertinence” (James 1890) the fact it has lingered so long testifies that it is not so easy to dispel. It may be false, but just like CI it is not obviously false.

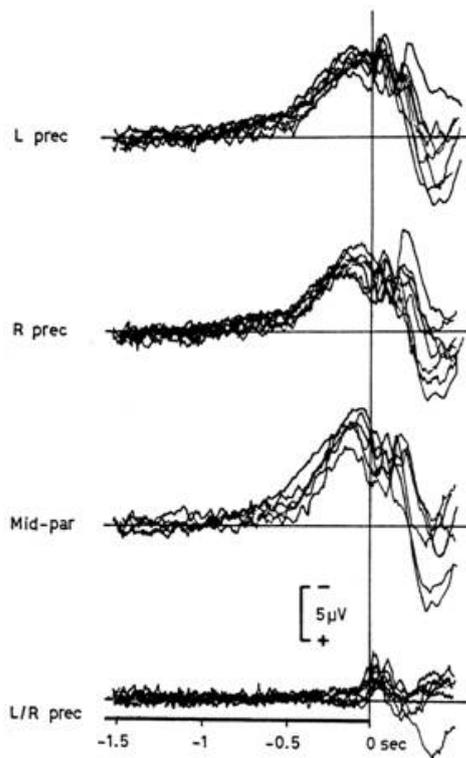
It is important to differentiate two varieties of EP (Revonsuo 2005). ‘Metaphysical EP’ is the view that consciousness is entirely without causal powers. Since having causal powers is central to at least one concept of what it is to be ‘real’ (Kim 1992), metaphysical EP may imply that consciousness is not real. ‘Biological EP’ is the view that consciousness may indeed have causal effects in the world but that these causal effects are not and never were involved in the reproductive success of conscious organisms. A good example is the thudding noise generated by the heart; this is indeed a causal effect of hearts but one that probably had nothing to do with selective advantage (see Section 3.1).

## **2.3 Empirical evidence for CI and EP**

Because neither EP nor CI can easily be dismissed nor proven on logical grounds alone, it is useful to consider the relevant experimental evidence. Experimental evidence in favor of CI would come from cases where behaviors for which consciousness has been assumed necessary are shown to be exhibited in the absence of consciousness. Experimental evidence in favor of EP would come from cases in which the causal link between consciousness and behavior is challenged.

One well-known experiment that seems at first blush to satisfy both the above criteria is the famous ‘readiness potential’ study of Benjamin Libet (Libet 1985), replicated and extended in various ways by Patrick Haggard and colleagues (Haggard 2005; Haggard

2008; Haggard & Eimer 1999). In the original Libet study, subjects were instructed to flex their right hand whenever they felt like it and as well to “pay close introspective attention to the instant of onset of the urge, desire, or decision to perform each such act and to the correlated position of a revolving spot on a clock face (indicating ‘clock time’)”. Throughout the experiment EEG potentials were recorded from the above the surface of the region of motor cortex implicated in hand movement. Strikingly, Libet found that the onset of easily recognisable motifs in the EEG traces (the ‘readiness potentials’, see Figure 1) reliably preceded awareness of the intention to act spontaneously, by around 350ms.



**Figure 1.** A typical recording of a readiness potential, as first identified by (Kornhuber & Deecke 1965), measured from four different regions of the cortical surface.

Libet’s findings have often been used to cast doubt on consciousness having a causal role in generating voluntary activity, thus providing evidence in favor of EP. As Libet asks, “if the brain can initiate a voluntary act before the appearance of a conscious intention ... is there any role for conscious function?”. Moreover, if apparently voluntary activity is determined prior to awareness of the intention to act, it seems plausible that this sort of action could in principle be carried out in the absence of consciousness, thus providing support for CI. But, even in the worst case that both of the above implications are justified, EP and CI will have been demonstrated only with respect to the specific sort of voluntary action investigated by Libet. From the statement that consciousness is inessential and/or epiphenomenal with regard to a certain sort of voluntary action it does not follow that consciousness is inessential and/or epiphenomenal in general.

Furthermore, as Libet notes, conscious causal powers are arguably preserved if consciousness can be assumed to have a ‘veto power’ over unconsciously initiated actions (free ‘won’t’ rather than free will). However, in this case there now arises the counter-problem of the possible existence of unconscious precursors to the conscious intention to veto an act. Moreover, there is an important methodological issue with the Libet experiment and other similar studies (Flanagan 1992): While the precise timing of subject hand flexions may be unpredictable, subjects are asked at the outset to make a conscious effort to let flexion occur spontaneously. Without having consciously assimilated the instructions it is unlikely that apparently spontaneous hand flexions would occur at all.

The ongoing brouhaha surrounding the Libet studies is troubling for many theorists. Why is it surprising that conscious processes are causally preceded by unconscious events? It seems that this could only be surprising given a Cartesian intuition that conscious intentions are ‘prime movers’ which are themselves without cause, or perhaps capable of being caused only by other conscious intentions (Mangan 2003). (Flanagan 1992) offers a neat summary: “Libet’s results, far from offering solace to the suspicious epiphenomenalist, are precisely the sort of results one would expect if one believes that conscious processes are subserved by nonconscious processes, and that conscious processes play variable but significant causal roles at various points in different cognitive domains.”

There is plenty of other evidence relevant to EP and CI. Studies in diverse fields including cognitive psychology, social psychology, and neuropsychology have shown that, contrary to our everyday intuitions, many complex human behaviors can occur in the absence of consciousness. In fact, some activities for which consciousness is commonly assumed to be useful can in fact be performed better unconsciously [(Dijksterhuis 2004), see Section 5.2]. While it is beyond the scope of the chapter to critically analyze all of this evidence, a good selection of it will appear throughout the remainder of this chapter.

## **2.4 Summary**

Although both CI and EP are likely to be false it is not easy to establish their falsity, whether by experiment or by theoretical argument. And even if CI and/or EP turn out to be true, we are still left with the deep questions of why conscious experiences appear to be uniquely associated with neural activity and not (as far as we know) with other physical processes (e.g., fermentation, photosynthesis), and why conscious contents almost always reflect functional roles [e.g., pains feel bad, sex feels good; see (Cole 2002)]. Nevertheless, both CI and EP serve as useful ‘null hypotheses’ to keep in mind as we discuss more positive proposals for the function(s) of consciousness. But before doing this we need a more detailed understanding of what is meant by ‘function’.

### 3.0 Functional analysis

To ask about the function of consciousness is to make at least two related inquiries: (i) Why did consciousness evolve? (ii) What does consciousness do?

#### 3.1 Why did consciousness evolve?

For many people an adequate functional explanation of a biological feature is one that plausibly accounts for its origin by natural selection. In this view, the function(s) of consciousness are the same as those features of consciousness that explain why it came to be present and maintained in certain organisms: The function of X is the effect that X has which explains why it is there. This interpretation of function can be called a 'teleological function', 'etioloical function', or 'proper function' (Millikan 1989): Following (Godfrey-Smith 1996) I will refer to this sort of function as a teleo-function.

Coming up with a plausible teleo-function for consciousness is challenging for a number of reasons. First, in complex highly interactive systems such as the brain it can be very difficult to make a direct connection from any effect of a part to a selective advantage enjoyed by the whole (Gould & Lewontin 1979). Second, the function a biological feature has in the present is not necessarily the function (if any) for which it was selected during evolutionary history. Our brains presumably were not selected for reading ability, yet reading is an important brain-dependent function for contemporary humans. Third, not all present-day biological features exist as a result of natural selection. For example, it is unlikely that the color of blood or the structure of the human chin were driven by natural selection (i.e., these features are not traits).<sup>3</sup> Although such biological EP may be more readily suspected for simple features than for apparently complex features such as consciousness (Grantham & Nichols 1999), we have seen in the foregoing that the EP suspicion is difficult to rule out.

A further problem lies in coming up with the right sort of evidence that can turn a 'how possibly' account into a 'how actually' account (Brandon 1995). That this is difficult for adaptationist explanations in general has led to the criticism that they are often little more than 'just so stories' (Gould & Lewontin 1979), although responses to this criticism have become increasingly vigorous over recent years [e.g., (Andrews et al. 2002)]. For consciousness in particular there is the additional hurdle of widespread skepticism that empirical evidence has anything to do with consciousness. This skepticism derives from the idea that consciousness, as a subjective phenomenon, does not directly engage with objective evidence. However, a science of consciousness requires only that we be epistemologically objective, a position which is quite consistent with the correct characterization of consciousness as ontologically subjective (Searle 1992).

According to (Brandon 1995), the sorts of evidence that are required for establishing teleo-functionality are: (i) evidence that selection has occurred (i.e., fossil evidence or

---

<sup>3</sup> Of course this does not exclude that the color of blood presently has functional significance for organisms, for example as an indicator of predatory threat or disease.

other experimental evidence); (ii) an ecological explanation of relative adaptedness; (iii) evidence that the traits in question are heritable; (iv) information about population structure; and (v) phylogenetic information about trait polarity (i.e., evidence that conscious organisms evolved from non-conscious organisms and not *vice-versa*). These criteria are not easy to satisfy (Polger in press). For example, fossil evidence for consciousness is difficult to imagine and the relevant experiments are hard to design and likely to be unethical; direct evidence for heritability is also hard to come by, and population structures in proposed adaptive environments for consciousness are mostly left unspecified. In short, coming up a solid adaptationist account of the evolution of consciousness is difficult and requires going well beyond establishing what consciousness *does* for an organism.

### **3.2 What does consciousness do?**

Instead of asking why consciousness evolved, we can ask instead what causal effects consciousness has with regard to present-day brains, bodies, and behaviors. In this view we are trying to isolate salient causal effects from among a multiplicity of effects that a given biological feature might have. A useful way to think about this is to consider the role played by the functionally characterized thing in how some larger system, of which the functionally characterized thing is a part, is able to exhibit a more complex capacity or behavior (Cummins 1975). For example, hearts have the function of pumping blood because this effect helps explain the capacity of the body to achieve circulation of oxygen. Following (Godfrey-Smith 1996) this sort of function can be called a ‘Cummins-function’ (another equivalent term is ‘causal role function’).

Of course in many cases, including the example just given, a teleological interpretation may be granted to the larger capacity (i.e., achieving oxygen circulation is likely to have been strongly selected for) but this interpretation does not necessarily transfer to the functionally characterized thing itself. In other words, characterizing a salient effect as a Cummins-function can avoid the mereological difficulties that sometimes attend ascriptions of teleo-functionality.

Both Cummins-functions and teleo-functions pick out functions from mere physical effects in virtue of explanatory salience (Godfrey-Smith 1996; Wright 1973); the thudding of the heart is neither teleo-functional nor Cummins-functional because it lacks explanatory salience. Also, proposing a Cummins-function for consciousness does not exclude a teleological role, it simply allows for the possibility that a readily identifiable teleological role may not exist. Therefore, on a Cummins-function view, we can evaluate the explanatory salience of consciousness with regard to the behavior of the organism as a whole without necessarily making the claim that consciousness was picked out by natural selection precisely in virtue of these explanatorily salient effects.

### **3.3 Functions and functionalism**

It is important to distinguish the question of the function(s) of consciousness from the philosophical position of functionalism, one of the most important developments in 20<sup>th</sup>

century analytic philosophy. The core thesis of functionalism is that mental states are second-order properties constituted by their causal relations to one another and to sensory inputs and motor outputs [(Fodor 1987; Putnam 1988), see also (Lewis 1972)]. For example, the mental state of being in pain is fully characterized by dispositions to say “ouch”, to wonder whether one is unwell, to take an aspirin, and so on. Functionalism remains a controversial position, most obviously because it implies that conscious states can be implemented in arbitrary physical systems and thus is seen by some to be too ‘liberal’ (Lycan 1987).

Within a functionalist framework the question of the function of consciousness is obviously important. However, accepting functionalism does not by itself negate CI or EP because causal relations need not always have explanatory salience (in the sense of being Cummins functions). Nor is it necessary to subscribe to functionalism in order to inquire about the functions of consciousness. It is conceivable that consciousness could have explanatorily salient causal effects in virtue of intrinsic properties of its material substrates, a ‘physicalist’ position that is contrary to functionalism.

### **3.4 Summary**

The function(s) of consciousness can be considered in terms of teleological significance or in terms of non-teleological explanatory salience. While it is tempting to look for adaptationist explanations for the function of consciousness, such explanations may be difficult to justify, as are adaptationist explanations in general.

## **4.0 Consciousness**

It is often pointed out that consciousness is not a unitary phenomenon (Block 2005; Zeman 2005). Asking about the function of consciousness requires clarifying the concept of consciousness itself.

### **4.1 Conscious level *versus* conscious content**

A first important distinction is between an organism being a conscious organism, and mental content being conscious mental content. A conscious organism is one which is capable of having conscious mental content. That is, a conscious organism has at any given time a particular level of consciousness. In humans, these levels range from complete unconsciousness (death, coma, general anesthesia) to full, awake and alert consciousness. Conscious content describes the continually changing phenomenal content (e.g., qualia such as redness and warmth) and intentional content (e.g., explicitly held beliefs) that is present for conscious organisms at non-zero conscious levels.<sup>4</sup>

---

<sup>4</sup> There are many other ways to carve these distinctions. For example, Rosenthal (2005) distinguishes being conscious *of* something which he calls “transitive consciousness”, from both “creature consciousness” (similar to conscious level), and “state consciousness”, which is the property of a mental state that makes it conscious.

Many types of mental content can be either conscious or unconscious. For example, during normal waking consciousness we can either hold the implicit (unconscious) belief that the sun will rise in the morning, or we can hold this belief explicitly (consciously). There is good evidence that the same is true for many other types of mental content. For example, masked priming experiments and the phenomenon of blindsight suggest that we can have unconscious perceptual content (Marcel 1983; Weiskrantz 1986).<sup>5</sup> It is also widely accepted that desires and emotions can occur unconsciously, although whether linguistic thoughts can exist without being conscious is less clear.

A consequence of the distinction between conscious level and conscious content is that the function that mental content has in virtue of being conscious cannot be inferred from the function of the organism's being conscious (Rosenthal 2008). In other words, the ability to have conscious content may serve a different (although probably overlapping) set of functions for an organism, than the fact that a particular sort of conscious content is present.

Where possible it is also useful to distinguish the function that mental content has in virtue of being conscious from the function that content would have even when unconscious (Rosenthal 2008). This can be difficult to do because a satisfactory distinction would require selectively carving off those mechanisms relevant to mental content being conscious mental content from those mechanisms underlying all other causal effects of that content. In a complex biological system it is not likely that such selective dismemberment will be possible. For example, phenomena such as blindsight may suggest that perceptual content can have function even when unconscious, however the visually guided behavior of a blindsight patient is usually worse than that of a healthy control [but see (Trevethan et al. 2007)]. Moreover, it can be argued that certain types of mental content, such as inner speech, cannot exist without being conscious.

#### **4.2 Primary consciousness *versus* higher-order consciousness**

A second key distinction is between primary consciousness and higher-order consciousness (Edelman 1989; Seth & Baars 2005).<sup>6</sup> Primary consciousness reflects the presence of a multimodal scene composed of perceptual and motor events. At its core, primary consciousness refers to the presence of a world; there is something like it is to be a primary conscious organism (Nagel 1974). Higher-order consciousness (HOC) reflects the observation that we (humans) are not only conscious, we are also conscious of being conscious. HOC involves the referral of the contents of primary consciousness to interpretative processes including a sense of self and, in more advanced forms, the ability to explicitly construct past and future scenes. While in humans these two forms of consciousness almost always go together (with possible exceptions in certain dreamlike or meditative states), it is conceivable that primary consciousness could exist

---

<sup>5</sup> Blindsight refers to the capability of some patients with visual cortex damage to perform visually guided behaviors even though they report the absence of any associated conscious content.

<sup>6</sup> Primary consciousness is sometimes referred to as 'sensory consciousness', and higher-order consciousness is sometimes referred to as 'reflective consciousness' or 'meta-consciousness'.

independently of HOC. Indeed, this may be case in many animals and perhaps as well in newborn infants (Seth et al. 2005).

The distinction between primary and higher-order consciousness is similar but not identical to that made by (Block 1995; Block 2005) between phenomenal consciousness (P-consciousness) and access consciousness (A-consciousness). Both HOC and A-consciousness involve metacognitive access, but HOC requires in addition that conscious contents are themselves higher-order whereas A-consciousness does not. Importantly, it is this shared feature of metacognitive access that allows us to verbally report conscious content and which therefore underwrites most experimental methods for studying consciousness, since it is only through explicit behavioral report that we (presently) are able to be sure whether a subject is conscious and if so know what she is conscious of. This also means that reported absence of conscious content could be due either to the absence of this content, or alternatively, to disruption of conscious access to this content that allows its report (Dehaene et al. 2006; Lamme 2006).

An influential philosophical framework that requires subtle interpretation in this context is David Rosenthal's higher-order thought (HOT) theory (Rosenthal 2005). According to HOT theory a mental state is a conscious mental state in virtue of the existence of a higher-order thought, distinct from that state, to the effect that one is in that state. That is, HOTs are not only essential for allowing report of conscious mental content but they are constitutive of that content, and any mental state can in principle occur without being conscious. HOTs themselves are rarely experienced consciously (i.e., having a HOT does not imply the existence of corresponding HOC content): the theory holds that to do so would require a corresponding third-order thought to the effect that one is having the second-order thought.

According to HOT theory, mental content has the same causal effects whether it is conscious or not. This is because the difference between consciousness and unconsciousness of mental content is due not to any property intrinsic to that content but instead is due to the presence or absence of the corresponding HOT. The function of consciousness therefore pertains to the content of these HOTs, which may well be distinct from that of their first-order target(s). As (Rosenthal 2008) argues, "[the] additional function that is due specifically to a first-order state's being conscious is simply the function of the higher-order state in virtue of which that first-order state is conscious. And it is natural to expect that the function of the higher-order state would be minimal relative to the function of the first-order state it is about."<sup>7</sup>

---

<sup>7</sup> By asserting that any mental state can occur without being conscious, HOT theory avoids an apparent obstacle for other theories in which some states cannot occur without being conscious. This obstacle exists to the extent that it is difficult to distinguish between the function a state has in virtue of being conscious from the function(s) it may have in virtue of other of its psychological or neurophysiological properties. However, if a theory can explain why a particular state entails that it is conscious (i.e., why it could not occur without being conscious), then an explanation of its function is also an explanation of its function in virtue of being conscious.

### **4.3 Summary**

There is a difference between an organism's being conscious, and the conscious contents that such an organism has. One set of functions cannot be inferred from the other. Much mental content can be either conscious or unconscious, and the function of conscious content ought, where possible, to be distinguished from the function that content would have independently of being conscious. Finally, consciousness is not a unitary phenomenon, differentiating most importantly into primary (sensory) and higher-order (metacognitive) varieties.

## **5.0 Volition and rationality**

We are now in a position to outline a number of possibilities for the functional roles consciousness may play in humans, and possibly as well in other species. We begin with some intuitively appealing ideas.

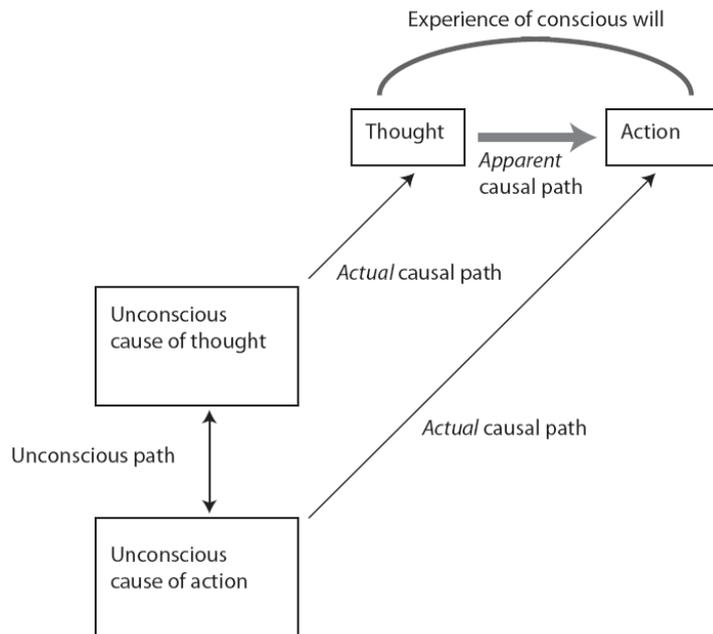
### **5.1 Volition**

The notion that function of consciousness is to initiate and control voluntary action has enormous appeal: We consciously think about doing X and then we do X. William James' 'ideomotor theory' closely follows this intuition by suggesting that actions are generated by having a thought about the action (James 1890). For example, getting out of bed on a chilly morning is caused by a conscious image of being out of bed accompanied by representations of the day's intended activities. Together, this mental content displaces any thoughts or images of staying-in-bed, and the getting-out-of-bed mental content is then translated into the appropriate motor commands.

A function for consciousness in volition is plausibly teleological and may correspond with both primary consciousness and HOC. However, the added value that volitional mental content has in virtue of being conscious is less clear. An alternative explanation of what happens on a chilly morning is that both the actual getting-out-of-bed and the volitional conscious experience of the intention-to-get-out-of-bed are caused by a common set of unconscious processes, perhaps including unconscious intentions. These unconscious processes can be attributed with volitional content precisely because they give rise to volitional conscious content as well as to actions that appear from an external perspective to be voluntary. On this view, the experience of volition is a conscious experience like any other and does not have any additional causal powers in virtue of its volitional content.

This interpretation has gathered support on both theoretical and empirical grounds. James himself left room for the possibility that voluntary actions may have unconscious causal precedents by saying that action-related thoughts have a tendency to cause the corresponding action; that is, a thought is not guaranteed always to produce the corresponding action. James, like Libet, also considered that conscious vetos ('acts of express fiat') could intervene to stop a given action from occurring (James 1890). More recently, Daniel Wegner has stated the strong position that conscious will is an illusion

and that we experience volition only when mental content is inferred, rightly or wrongly, to have produced the corresponding physical action (Wegner 2002). According to Wegner's theory of 'apparent mental causation' we tend to make inferences resulting in an experience of volition only when the corresponding mental content satisfies the constraints of primacy (the content immediately preceded the action), consistency (the content corresponds to the action), and exclusivity (there is no other plausible causal factor) [(Wegner 2003), see figure 2].



**Figure 2.** Wegner (2003) distinguishes two causal pathways, one leading from unconscious causes of actions to actions and another leading from unconscious action-related mental content to conscious experience of that mental content. That conscious mental content becomes an experience of volitional mental content when an inference is made of an apparent causal pathway from the mental content to the action, which will happen when the constraints of primacy, consistency, and exclusivity are satisfied. Figure adapted from Wegner (2003).

If conscious will is illusory it should be possible for voluntary actions to occur in the absence of experiences of volition. Several neurological and psychological disorders show these features. For example, patients with 'alien hand syndrome' report that one of their hands seems to be outside of their voluntary control, performing apparently voluntary complex action sequences without any accompanying experience of voluntary control (Geschwind et al. 1995). Schizophrenic auditory hallucinations also produce anomalous experiences of will in which the patients' own volitional thoughts are attributed to others, or even to television sets (Hoffman 1991).

Experiments in normal subjects have also addressed the issue of volition. On one hand the Libet studies attempt to show that voluntary actions do *not* involve consciousness

(Section 2.3). On the other, several lines of evidence suggest that consciousness *is* important for controlling voluntary actions, but as with the Libet studies, their interpretation is rarely straightforward. In Jacoby's 'exclusion task' (Jacoby 1991) subjects are presented with a target word (e.g., 'reason') and then with the stem of a word (e.g., 'rea'). The task is to complete the stem into any word except the target (e.g., 'realize'). When subjects report conscious perception of the target they are usually able to perform this task successfully, but when they do not report conscious perception of the target (e.g., if it is presented very rapidly and/or followed by a masking stimulus) then subjects tend to fail, erroneously completing the word-stem with the target. These results are often taken to show a link between voluntary action and consciousness, since subjects are only successful at excluding words that they were conscious of having seen. But subjects who have not consciously seen the word are conscious of having not seen it; therefore, completing the word according to the not-consciously-seen (but still perceptually processed) word is still consistent with the task instructions (Rosenthal 2008). Also, studies with hypnotic subjects amply demonstrate that many types of exclusion can occur in the absence of reportable conscious experience of the excluded target [e.g., (Evans 1980); see (Dienes & Perner 2007) for a review].

Perhaps more compelling are experiments showing that blindsight subjects are unable to initiate voluntary actions with respect to blind-field stimuli (Marcel 1983) or to suppress corrective movements during rapid pointing [(Pisella et al. 2000) see (Danckert & Rossetti 2005) for a review]. Conversely, normally unconscious actions can be brought under voluntary control by biofeedback training, but only when subjects are given conscious feedback of the process (Roger & Galand 1981). In both cases, the essential link seems to be that consciousness of a stimulus is necessary for both voluntary actions, and for the conscious experience of volition with respect to that stimulus. But in neither case does it follow that the conscious experience of volition causes the voluntary action.

## **5.2 Rational action**

The association of consciousness with rationality is equally intuitive as the association with voluntary action. Introspection suggests strongly that conscious deliberation can supply rational responses in situations in which unconscious, automatic reactions may fail, and philosophers and scientists from Descartes and Locke onwards have emphasized the benefits of conscious deliberation in decision making. Block, for instance, holds that mental content is access conscious (A-conscious) if it is "poised for use as a premise in reasoning, ... [and] for [the] rational control of action and speech" (Block, 1995, p.231). Cognitive and neural theories of consciousness that stress global access and integration of conscious content are also aligned with rationality as a function since on this view conscious contents will be accessible to a wide range of potentially rational cognitive processes (Section 6). Along similar lines Koch argues that the function of consciousness is to provide an "executive summary" to those parts of the brain involved in planning and deliberation (Crick & Koch 1995; Koch 2004).

A rational action function for consciousness seems to associate more with HOC than with primary consciousness, and as with volition is plausibly teleological. But again, empirical

evidence weighs against there being a direct connection: Not all conscious thinking is rational, and not all rational behavior is conscious. The dissociation between rational thinking and rational acting is strikingly illustrated by episodes in which subjects provide false (confabulatory) rationalizations for the causes of their actions. The classic experiments of (Nisbett & Wilson 1977) showed that people's descriptions of their own reasoning processes are often inaccurate and seem to be modeled after logical-sounding idealized reasoning processes rather than the process actually used: "It is often the result of a reasoning process, rather than the reasoning process itself which shows up in conscious thought" (Nisbett & Wilson 1977).<sup>8</sup>

The recently developed 'choice blindness' paradigm extends Nisbett and Wilson's results in a novel direction (Johansson et al. 2006). Subjects are shown pairs of pictures of female faces, asked to select the most attractive and then to describe why they made each selection. Unknown to the subjects, on some trials the pictures are covertly switched, with the switch carried out *after* a choice was made but *before* introspective feedback is sought. Remarkably, on some of these 'switch' trials the subjects fail to detect the switch but nevertheless offer a plausible account why they chose a particular face, even though that had actually selected the other face. As with Nisbett and Wilson, these results show a clear divergence between conscious thinking and rational action.

The association between consciousness and rationality is further challenged by results showing that certain decisions can be *more* rational when subjects are *not* consciously aware of the decision making process. For example, (Dijksterhuis 2004; Dijksterhuis et al. 2006) hypothesized – based on the limited capacity of consciousness - that unconscious deciding could have an advantage over conscious deciding in complex situations involving tradeoffs among multiple factors. In support of this hypothesis they found that subjects deciding among complex alternatives (e.g., different makes of car) performed worse when they were encouraged to mull consciously over the different options, than when they were prevented from engaging in conscious deliberation. This performance difference was apparent both by objective criteria (i.e., when there really was a 'best' car) and by subjective criteria (i.e., post-decision satisfaction), and it disappeared when the decision problem was simple (e.g., different types of towel).

These results resonate with neurological evidence that an excess of rationality can be maladaptive. (Damasio 1994) describes numerous examples of patients with damage to prefrontal cortex who are hampered in their everyday life by an inability to make decisions effectively and efficiently. In one example Damasio describes a patient who endlessly enumerated the pros and cons of two suggested appointment dates, apparently paying as much attention to this trivial task as might be appropriate for deciding between two mortgages. Damasio's suggestion is that normal consciousness aids rational decision making not by facilitating rational thinking *per se*, but by biasing rational deciding in certain ways in order to reduce the space of possible options and the time and effort required to decide among them. This bias and channeling occurs as a result of the

---

<sup>8</sup> Nisbett and Wilson's results have often been falsely interpreted as undermining the scientific validity of introspection. On the contrary, their results show an interesting divergence between subjects' experience and their actions.

integration of emotional valence into conscious content related to the decision options (Damasio 1994; Damasio 2000).<sup>9</sup>

### **5.3 Summary**

Consciousness should not be excluded from functional roles in volition and rationality. Consciousness may be necessary for many aspects of volition; conscious decision making and conscious actions in general are often rational, and in many cases conscious deliberation may serve rational ends. But it is not the experience of volition that causes the voluntary action and a wealth of experiments show that consciousness is neither necessary nor sufficient for rational action.

## **6.0 Flexible action and the integration consensus**

A recurring idea in theories of consciousness is that consciousness serves to integrate otherwise independent neural and cognitive processes. This ‘integration consensus’, which can be traced back at least to (Sherrington 1906), has been expressed most forcefully in the cognitive context by Baars’ ‘global workspace’ theory (Baars 1988; Baars 2002) and in the neurophysiological context by Edelman (Edelman 1989; Edelman 2003; Edelman & Tononi 2000) and Tononi (Tononi 2004; Tononi & Edelman 1998).<sup>10</sup> Most participants within the integration consensus see consciousness as a “supremely functional adaptation” (Baars 1988), particularly with respect to enabling flexible, context-dependent behavior. As we will see below, functional proposals based on the integration consensus are explicitly teleological and associate with both primary consciousness and with HOC.

### **6.1 Global workspace theory**

The cornerstone of global workspace (GW) theory is that consciousness involves a central resource (the GW) which enables distribution of signals among numerous otherwise informationally encapsulated and functionally independent specialized processors (Baars 1988; Baars 2002). GW theory states that mental content becomes conscious mental content when it gains access to the GW such that it can influence a large part of the brain and a correspondingly wide range of behaviors (in this sense GW theory embodies Block’s notion of A-consciousness). This principle invites a televisual or theatrical metaphor: A process becomes conscious when it is broadcast widely, or when it comes on stage in the presence of a large audience, but not when it remains private. A key aspect of GW theory is that conscious contents unfold in an integrated, serial manner but are the product of massively parallel activity among specialized processors. The integrated states of the GW follow each other in a meaningful but

---

<sup>9</sup> (Cabanac 1996) also identifies the primary function of consciousness in affect, proposing that conscious pain and pleasure serve as common currencies, obviating the need for organisms to store large numbers behavioral ‘rules of thumb’ and thereby contributing to behavioral flexibility (see Section 6). The notion of common currency in this context is explicitly teleological.

<sup>10</sup> The term ‘integration consensus’ is due to Morsella (2005).

complex progression that depends on multiple separate processes, each of which might have something of value to add to the ongoing constitution of the GW.

According to Baars the premier function of conscious integration is to provide behavioral flexibility. There is good evidence that unconscious processes can be extremely rapid and free from the sorts of capacity limits that attend conscious processing but that they are also more fixed and predetermined (Andersen 1983; Schneider et al. 1994). GW theory accounts for this evidence by suggesting that in familiar situations automatic unconscious processors reel off 'canned' responses with high efficiency, but in novel situations the broadcast of multimodal signals within the GW can mediate the production of novel responses (Baars 1988).

The functional role of consciousness within GW theory has been refined in several directions. (Baars 1988) notes that the interplay of serial and parallel processes in a GW architecture can subserve analogical reasoning because the integrated content of the GW can stimulate a wide range of unconscious processes to locate analogical content. (Shanahan & Baars 2005) further argue that this process can provide a solution to the notorious 'frame problem', i.e. the problem of dealing effectively with signals from potentially any domain without having to explicitly sift the relevant from the irrelevant (Dennett 1978; Fodor 1983). This argument is based on the reasoning that any solution to the frame problem must involve an informationally unencapsulated process, i.e., one that can draw on information from potentially anywhere (Fodor 2000), and that analogy formation exemplifies information unencapsulation. By the same token, an informationally unencapsulated architecture can underwrite behavioral flexibility by allowing the effective and efficient integration of multiple cognitive processes to produce something new.

A different and explicitly teleological take on the functional utility of a GW architecture is that it allows an organism to rely more on mental simulation and internal evaluation to select actions, reducing both energy expenditure and risk (Dehaene & Naccache 2001). These authors argue that the GW, by allowing a wide range of cognitive processes to bear on action selection, participates in an evolutionary trend towards increasing internalization of environmental representations whose main advantage is the freeing of the organism from its immediate environment. This theme has been pursued further by (Hesslow 2002) and (Revonsuo 2005) who argue that internal simulation of behavior and perception can explain the appearance in consciousness of an 'inner world', and by (Shanahan 2006) who has incorporated an internal simulation loop into a computational model of a GW architecture.

As with most other participants in the integration consensus, GW theory is vulnerable to both EP and CI as emphasized by computational models that embody the principles and functionality of the GW architecture but for which the attribution of consciousness seems implausible (Dehaene et al. 2003; Franklin & Graesser 1999; Shanahan 2006); similarly, on HOT theory it may be that the function provided by access to the GW may be independent of the state being a conscious state (Rosenthal, 2008). Also, integration for flexibility is arguably a very basic function of nervous system activity that is substantially

present even in invertebrates (Ferguson & Benjamin 1991a; Ferguson & Benjamin 1991b), although whether a GW architecture is present in these cases remains an open question.

## **6.2 Skill acquisition and learning**

Flexible control is needed especially during acquisition of new skills. Many behavioral observations have indicated that acquiring a new skill requires conscious attention during initial phases, but that as learning progresses the execution of the skill becomes increasingly automatic (Schneider et al. 1994; Schneider & Shiffrin 1977). Consistent with this result, a recent fMRI study of motor sequence learning showed a shift from widespread cortical involvement during early learning to predominantly subcortical activation during later learning phases (Floyer-Lea & Matthews 2004). However even expert behavior is still accompanied by conscious content, though it is usually sensory and kinesthetic rather than verbal or thoughtlike (Dreyfus & Dreyfus 1982; Merleau-Ponty 1962). Therefore any functional role for consciousness in skill acquisition is likely to correspond more to HOC than to primary consciousness.

Victor Lamme has argued more generally in favour of a learning function for consciousness. According to Lamme, consciousness arises from the interaction of a 'feedforward sweep' of stimulus-evoked neural activity and a 'recurrent' or 'reentrant' sweep originating in frontoparietal areas (Lamme & Roelfsema 2000), an idea that fits neatly within the integration consensus. Lamme argues that the recurrent sweep promotes synaptic plasticity (and hence learning) by allowing pre- and post-synaptic neurons to be active simultaneously. Challenge for this view include ample evidence for unconscious learning, both in human subjects (Stadler & Frensch 1997) and in machines, as well as theoretical difficulties in establishing direct connections between synaptic plasticity and behavioral learning (Seth & Edelman 2007).

## **6.3 The boundaries of integration**

An important issue within the integration consensus is, Which kinds of signals are capable of being integrated into conscious scenes and which are not? For example, neural activity related to vegetative functions and to low-level perceptual processes does not evoke conscious contents (Laureys 2005). A useful approach this question is to look for common features when contrasting conscious and unconscious processes (Baars 1988; Baars 2002).

According to (Morsella 2005)'s 'supramodular interaction theory' conscious contents mediate interactions among 'supramodular response systems'. These are systems which have different high-level concerns (e.g., the food intake system, the instrumental learning system) and which can come into conflict at the level of the skeleomotor system. Morsella points out that consciously impenetrable processes such as pupillary reflexes, peristalsis, and bronchial dilation do not involve control of skeletal muscle. By contrast, consciously penetrable processes like inhaling, coughing, swallowing, and defecating all do. On this view, the function of consciousness is to mediate interactive processing

across subsystems to allow the organism at any given time to produce a single, adaptive, skeleomotor action.

A different contrastive approach is taken by (Merker 2005), whose starting point is the stability of the consciously perceived world. Merker emphasizes the remarkable nature of this stability given the confounding influence of self-produced motion of sensor arrays mounted on multijointed and swivelling body parts. Conscious contents successfully exclude both the multiple sensory and sensorimotor transformations that are needed to extract a stable world image, and the complex coordinations of muscle movements needed to produce actions. What is left is a stable arena for decision-making and for planning our actions, and the function of consciousness in this view is precisely to provide and maintain this stable arena. An interesting implication of this proposal is that consciousness will be present in all organisms which face similar problems of coordination and neuronal logistics.

#### **6.4 Discrimination and complexity**

In an influential paper (Tononi & Edelman 1998) observed that conscious scenes are not only integrated, they are also at the same time differentiated from each other [see also (Edelman 2003; Edelman & Tononi 2000; Seth et al. 2006)]. Not only is every conscious scene experienced ‘all of a piece’ (integration) but every conscious scene is also unique (differentiation). Thus the occurrence of any conscious scene simultaneously rules out the occurrence of a vast number of alternative conscious scenes. In the strict sense that information corresponds to a reduction in uncertainty, every conscious scene, primary or higher-order, is therefore enormously informative. This is functional for the organism because each differentiated conscious scene can be linked to a different behavioral response. On this view, the function of consciousness is adaptive and flexible discrimination.

The foregoing is at the heart of the ‘dynamic core hypothesis’ [DCH, (Edelman & Tononi 2000; Tononi & Edelman 1998)], which is part of the more general theoretical framework provided by the ‘theory of neuronal group selection’ [TNGS, (Edelman 1987; Edelman 1989; Edelman 2003)]. According to the DCH, the balance between differentiation and integration in every conscious scene is underpinned by a corresponding balance in the neural dynamics responsible for consciousness, which in turn implies that consciousness is generated by interactions among widely distributed groups of neurons. A key feature of the DCH is the proposal of a quantitative measure for this balance, ‘neural complexity’, which uses information theory to express the extent to which large subsets of a system tend to behave coherently and small subsets tend to behave independently (Tononi et al. 1994) [I have recently proposed an alternative quantitative measure, based on multivariate autoregressive modeling, called ‘causal density’ (Seth 2008; Seth et al. 2008)]. Computational modelling has shown that the highly reentrant neuroanatomy of the thalamocortical system is particularly well suited to producing dynamics of high neural complexity, whereas other neural systems such as the cerebellum and basal ganglia are not (Sporns et al. 2000). Accordingly, the DCH proposes that the neural mechanisms

underlying conscious experience consist of a functional cluster in the thalamocortical system, this being the dynamic core.

A recent variant of the DCH, the ‘information integration theory’ (IITC), proposes a different quantitative measure,  $\Phi$ , which is based on identifying the informational ‘weakest link’ within a system (Tononi 2004). Whereas the DCH proposes that high values of neural complexity may be necessary but not sufficient for consciousness, the IITC proposes that  $\Phi$  is by itself an adequate measure of the quantity of consciousness generated by a system [see (Seth et al. 2006) for further comparison]. Therefore, according to the IITC, the function of consciousness is to integrate information for the simple reason that conscious experience is defined as information integration.

The DCH and the IITC sharpen and quantify the integration consensus and relate it directly to neurophysiological processes. They attempt to escape both CI and EP by explicitly identifying consciousness with discrimination (DCH) or information integration (IITC). However, this move can be criticized as ‘defining away’ the problem (Chalmers 1996) and in addition it may be possible to find examples of unconscious systems with arbitrarily high neural complexity and  $\Phi$  (Seth et al. 2006). Other neurophysiological perspectives that participate in the integration consensus include so-called ‘field’ theories of consciousness [e.g., (John 2001; Kinsbourne 1988)], but these theories make similar claims regarding conscious function and will not be discussed further here.

## **6.5 Neuronal group selection**

While the DCH derives from the TNGS, the latter offers additional perspectives on the function of consciousness that extend beyond information integration and discrimination among conscious scenes.

The TNGS (also known as ‘neural Darwinism’) is a large-scale biological theory of brain function that has roots in evolutionary biology and immunology in viewing brain operations as selectionist in nature, rather than instructionist, like a computer (Edelman 1987). According to the TNGS, primary consciousness reflects an adaptive linkage of current perceptual categorization to past learning responses and to future needs. This linkage is value-dependent, where ‘value’ reflects salience to the organism as mediated by neuromodulatory systems (e.g., the dopaminergic, cholinergic and noradrenergic systems). The result of these interactions is the generation of a ‘remembered present’, a description that evokes James’ ‘specious present’ (James 1890) by emphasizing the historically contextualized nature of ongoing primary conscious experience. According to the TNGS, organisms in possession of a remembered present will enjoy increased discriminatory selectivity, flexibility, and planning capacity when responding to complex environments, as compared to their preconscious and unconscious competitors and ancestors. The added value of HOC in the TNGS is that the dependence of consciousness on present inputs is no longer limiting (Edelman 2003). The ability explicitly to construct past and future scenes extends the integrative capacity of consciousness, allowing the

development and deployment of more sophisticated, flexible, and adaptive actions and action plans.

## **6.6 Summary**

According to the integration consensus, consciousness functions to bring together diverse signals in the service of enhanced behavioral flexibility and discriminatory capacity. Theoretical proposals within this consensus are among the most highly developed and are increasingly open to experimental testing. However, integration theories must explain *why* consciousness is necessary since many integrative functions seem plausibly executable by unconscious devices. The DCH and the IITC address this issue by relating phenomenology and complexity, but for these theories it remains unclear whether high values of neural complexity (or  $\Phi$ , or causal density) are sufficient for consciousness.

## **7.0 Beyond the integration consensus: Alternative functions**

To finish we discuss several alternative ideas which both compete with and complement integrative functions. Because these proposals tend to associate consciousness with one or more existing cognitive functions they are, as usual, vulnerable to both EP and CI.

### **7.1 Error correction**

Rodolfo Llinas has argued that the most general of all brain functions is prediction (Llinas 2004) [or reduction of prediction error, see FRISTON REF]. Predictions help an animal anticipate appetitive and aversive events and facilitate the formulation and execution of the appropriate motor responses. Much of this prediction happens unconsciously. For example, when driving a familiar route we often arrive without remembering much about the journey itself. However, if during the journey another car unexpectedly swerves in front of us our conscious contents suddenly become dominated by the experience of driving. This example illustrates the idea that consciousness functions to detect and allow correction of prediction errors during behavior [(Gray 2004; Pally 2005), but see (Pisella et al. 2000)]. According to the ‘comparator hypothesis’ (Gray 2004), consciousness is part of the brain’s monitoring of whether its expectations have occurred or have failed to occur: “What we experience consciously has been selected for its value on a scale of expected versus unexpected” (Gray, 2004, p.232). Because error correction involves metacognitive monitoring of ongoing mental content, any functional role for consciousness in detecting and correcting prediction errors would apply more to HOC than to primary consciousness. After all, when driving a familiar route we are not wholly unconscious, rather, we may be unaware that we are having the experience of driving.

### **7.3 Social interactions**

Day-to-day human existence involves negotiating a maze of social interactions which in turn conceiving of other people as beings with minds. According to (Humphrey 1982; Humphrey 2002), we humans must be excellent “natural psychologists”, quickly and

effortlessly attributing to other people mental content such as beliefs, desires, moods, sensations, and the like. Humphrey's proposal is that consciousness, in particular HOC, fulfils this functional role: By being conscious of our own mental content we acquire an enhanced ability to infer the mental content of others, especially those that belong to our own social group. This view challenges the attribution of consciousness to creatures that lack highly developed mechanisms of metacognitive access and/or intricate social lives; such creatures include many non-human animals and possibly also infants and autists.

#### **7.4 Dreaming**

Most discussions of the function of consciousness focus on conscious contents during wakefulness, but much of our lifetime conscious experience occurs during sleep (i.e., when dreaming) and dream content is substantially different from waking content. For example, dreams have the property of naïve-realism and certain qualia, such as odors, are rarely present (Metzinger 2003).

Dreams pose an apparent challenge for ascriptions of function because the dreaming brain is actively inhibited from generating behavior. But the fact that behavior is restricted or absent during dreaming does not mean that dream content is causally impotent. It is easy to imagine that dream content can have causal power by affecting future behavior in subsequent waking states. (Revonsuo 2005) has proposed that consciousness in general has the function of providing a virtual reality arena in which the consequences of actions can be tried out without taking the action itself (Hesslow 2002), and that this function is nowhere more evident than in dreaming. Dream content is often disturbing, threatening, and is associated with anxiety much more than waking consciousness, even in subjects without anxiety disorders or depression. According to Revonsuo, dreaming was originally a biological defense mechanism for simulating threat perception and rehearsing threat avoidance responses.

#### **7.5 Artificial consciousness**

Artificial consciousness (AC, equally, 'machine consciousness') refers to attempts to model and/or create consciousness in machines (Chella & Manzotti 2007). AC has the potential to sharpen theories of conscious function because AC models are developed explicitly to perform some function and are analyzed according to their ability to do so (Clowes & Seth 2008). Also, model construction allows the limitations of theoretical arguments to be identified and extended, for example in the enriched view of 'virtual machine functionalism' (Sloman & Chrisley 2003) according to which functional properties arise from interactions among machines as well among mental states *per se*. Many current AC models stress *control* functions, for example (Sanz et al. 2007) suggest that consciousness may reflect the operation of a "model-based reflective predictive controller".

## 8.0 Conclusions

While there may always remain suspicious epiphenomenalists and die-hard conscious inessentialists, there is abundant and increasing evidence that consciousness is functional. This evidence pertains both to the functional utility of being a conscious organism, and to having particular conscious content. According to the integration consensus, being a conscious organism allows for the adaptive integration of many input and output signals in the service of behavioral flexibility, and the particular conscious content that is integrated functions to elicit a particular adaptive response. But because consciousness is a constellation concept covering a range of possible distinguishable processes, future experiments and theoretical developments will doubtless refine and differentiate the range of conscious functions beyond those discussed here.

## Acknowledgements

I am grateful to Bernard Baars, Zoltan Dienes, Bjorn Merker, Thomas Metzinger, and especially David Rosenthal for their helpful comments on a first draft of this chapter.

## References

- Andersen, J. 1983 *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Andrews, P. W., Gangestad, S. W. & Matthews, D. 2002 Adaptationism--how to carry out an exaptationist program. *Behav Brain Sci* **25**, 489-504; discussion 504-53.
- Baars, B. J. 1988 *A cognitive theory of consciousness*. New York, NY: Cambridge University Press.
- Baars, B. J. 2002 The conscious access hypothesis: origins and recent evidence. *Trends Cogn Sci* **6**, 47-52.
- Block, N. 1995 On a confusion about a function of consciousness. *Behavioral and Brain Sciences* **18**, 227-247.
- Block, N. 2005 Two neural correlates of consciousness. *Trends Cogn Sci* **9**, 46-52.
- Brandon, R. N. 1995 *Adaptation and Environment*. Princeton, NJ: Princeton University Press.
- Cabanac, M. 1996 On the origin of consciousness: A postulate and its corollary. *Neurosci Biobehav Rev* **20**, 33-40.
- Chalmers, D. J. 1996 *The conscious mind: In search of a fundamental theory*. New York, NY: Oxford University Press.
- Chella, A. & Manzotti, R. (ed.) 2007 *Artificial consciousness*. Exeter: Imprint Academic.
- Clowes, R. W. & Seth, A. K. 2008 Axioms, properties and criteria: Roles for synthesis in the science of consciousness. *Artif Intell Med*.
- Cole, D. 2002 The functions of consciousness. In *Consciousness Evolving* (ed. J. Fetzer). Philadelphia: John Benjamins.
- Crick, F. & Koch, C. 1995 Are we aware of neural activity in primary visual cortex? *Nature* **375**, 121-3.
- Cummins, R. 1975 Functional analysis. In *Journal of Philosophy*, vol. 72, pp. 741-764.
- Damasio, A. 1994 *Descartes' error*. London: MacMillan.

- Damasio, A. 2000 *The feeling of what happens: Body and emotion in the making of consciousness*: Harvest Books.
- Danckert, J. & Rossetti, Y. 2005 Blindsight in action: what can the different sub-types of blindsight tell us about the control of visually guided actions? *Neurosci Biobehav Rev* **29**, 1035-46.
- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J. & Sergent, C. 2006 Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci* **10**, 204-11.
- Dehaene, S. & Naccache, L. 2001 Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1-37.
- Dehaene, S., Sergent, C. & Changeux, J. P. 2003 A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci U S A* **100**, 8520-5.
- Dennett, D. 1978 *Brainstorms*. Cambridge, MA: MIT Press.
- Dennett, D. 1991 *Consciousness Explained*. Boston, MA: Little, Brown, and London.
- Dienes, Z. & Perner, J. 2007 The cold control theory of hypnosis. In *Hypnosis and conscious states: The cognitive neuroscience perspective* (ed. G. Jamieson), pp. 293-314. Oxford: Oxford University Press.
- Dijksterhuis, A. 2004 Think different: The merits of unconscious thought in preference development and decision making. *Journal of Personality & Social Psychology* **87**, 586-598.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & van Baaren, R. B. 2006 On making the right choice: the deliberation-without-attention effect. *Science* **311**, 1005-7.
- Dretske, F. 1997 What good is consciousness? *Canadian Journal of Philosophy* **27**, 1-17.
- Dreyfus, H. & Dreyfus, S. 1982 *Mind over machine*: Free Press.
- Edelman, G. M. 1987 *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books, Inc.
- Edelman, G. M. 1989 *The remembered present*. New York, NY: Basic Books.
- Edelman, G. M. 2003 Naturalizing consciousness: a theoretical framework. *Proc Natl Acad Sci U S A* **100**, 5520-4.
- Edelman, G. M. & Gally, J. 2001 Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. USA* **98**, 13763-13768.
- Edelman, G. M. & Tononi, G. 2000 *A universe of consciousness : how matter becomes imagination*. New York, NY: Basic Books.
- Evans, F. J. 1980 Posthypnotic amnesia. In *Handbook of hypnosis and psychosomatic medicine* (ed. G. D. Burrows & L. Dennerstein), pp. 85-103. Amsterdam: Elsevier.
- Ferguson, G. P. & Benjamin, P. R. 1991a The whole-body withdrawal response of *Lymnaea stagnalis*. I. Identification of central motoneurons and muscles. *J Exp Biol* **158**, 63-95.
- Ferguson, G. P. & Benjamin, P. R. 1991b The whole-body withdrawal response of *Lymnaea stagnalis*. II. Activation of central motoneurons and muscles by sensory input. *J Exp Biol* **158**, 97-116.
- Flanagan, O. 1992 *Consciousness Reconsidered*. Cambridge, MA: MIT Press.
- Floyer-Lea, A. & Matthews, P. M. 2004 Changing brain networks for visuomotor control with increased movement automaticity. *J Neurophysiol* **92**, 2405-12.

- Fodor, J. A. 1983 *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. A. 1987 *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. A. 2000 *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.
- Franklin, S. & Graesser, A. 1999 A software agent model of consciousness. *Conscious Cogn* **8**, 285-301.
- Geschwind, D. H., Iacono, M., Mega, M. S., Zaidel, D. W., Cloughesy, T. & Zaidel, E. 1995 Alien hand syndrome: interhemispheric motor disconnection due to a lesion in the midbody of the corpus callosum. *Neurology* **45**, 802-8.
- Godfrey-Smith, P. G. 1996 *Complexity and the function of mind in nature*. Cambridge, MA: MIT Press.
- Gould, S. J. & Lewontin, R. C. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* **205**, 581-98.
- Grantham, T. & Nichols, S. 1999 Evolutionary psychology: Ultimate explanations and Panglossian predictions. In *Where biology meets psychology: Philosophical essays* (ed. V. Hardcastle), pp. 47-67. Cambridge, MA: MIT Press.
- Gray, J. 2004 *Consciousness: Creeping up on the hard problem*. Oxford: Oxford University Press.
- Haggard, P. 2005 Conscious intention and motor cognition. *Trends Cogn Sci* **9**, 290-5.
- Haggard, P. 2008 Human volition: towards a neuroscience of will. *Nat Rev Neurosci* **9**, 934-946.
- Haggard, P. & Eimer, M. 1999 On the relation between brain potentials and the awareness of voluntary movements. *Exp Brain Res* **126**, 128-33.
- Hesslow, G. 2002 Conscious thought as simulation of behaviour and perception. *Trends Cogn Sci* **6**, 242-247.
- Hoffman, R. E. 1991 The Duphar Lecture: on the etiology of alien, nonself attributes of schizophrenic 'voices'. *Psychopathology* **24**, 347-55.
- Humphrey, N. 1982 *Consciousness regained*. Oxford: Oxford University Press.
- Humphrey, N. 2002 The uses of consciousness. In *The mind made flesh: Consciousness and the physical world* (ed. N. Humphrey), pp. 65-85. Oxford: Oxford University Press.
- Jacoby, L. 1991 A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language* **30**, 513-541.
- James, W. 1890 *The principles of psychology*. New York: Henry Holt.
- Johansson, P., Hall, L., Sikstrom, S., Tarning, B. & Lind, A. 2006 How something can be said about telling more than we can know: on choice blindness and introspection. *Conscious Cogn* **15**, 673-92; discussion 693-9.
- John, E. R. 2001 A field theory of consciousness. *Conscious Cogn* **10**, 184-213.
- Kim, J. 1992 'Downward causation' in emergentism and nonreductive physicalism. In *Emergence of reduction? Essays on the prospect of nonreductive physicalism* (ed. A. Beckermann, H. Flohr & J. Kim), pp. 119-138. Berlin: de Gruyter.
- Kinsbourne, M. 1988 Integrated field theory of consciousness. In *Consciousness in contemporary science* (ed. A. J. Marcel & E. Bisiach), pp. 239-256. New York: Oxford University Press.

- Kirk, R. 1974 Zombies vs materialists. *Proceedings of the Aristotelian Society, Supplementary Volume* **48**, 135-52.
- Koch, C. 2004 *The Quest for Consciousness: A Neurobiological Approach*: Roberts and co.
- Kornhuber, H. H. & Deecke, L. 1965 [Changes in the Brain Potential in Voluntary Movements and Passive Movements in Man: Readiness Potential and Reafferent Potentials.]. *Pflugers Arch Gesamte Physiol Menschen Tiere* **284**, 1-17.
- Lamme, V. 2006 Towards a true neural stance on consciousness. *Trends Cogn Sci* **10**, 494-501.
- Lamme, V. A. & Roelfsema, P. R. 2000 The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* **23**, 571-9.
- Laureys, S. 2005 The neural correlate of (un)awareness: lessons from the vegetative state. *Trends Cogn Sci* **9**, 556-9.
- Leonardo, A. 2005 Degenerate coding in neural systems. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* **191**, 995-1010.
- Lewis, D. 1972 Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* **50**.
- Libet, B. 1985 Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* **8**, 529-566.
- Llinas, R. R. 2004 *I of the vortex*. Cambridge, MA: MIT Press.
- Lycan, W. G. 1987 *Consciousness*. Cambridge, MA: MIT Press.
- Mangan, B. 2003 Volition and property dualism. *Journal of Consciousness Studies* **10**, 29-34.
- Marcel, A. J. 1983 Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology* **15**, 197-237.
- Merker, B. 2005 The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. *Conscious Cogn* **14**, 89-114.
- Merleau-Ponty, M. 1962 *Phenomenology of perception*: Routledge & Kegan Paul.
- Metzinger, T. 2003 *Being No-One*. Cambridge, MA: MIT Press.
- Millikan, R. 1989 In defense of proper functions. *Philosophy of Science* **56**.
- Moody, T. 1994 Conversations with zombies. *Journal of Consciousness Studies* **1**, 196-200.
- Morsella, E. 2005 The function of phenomenal states: supramodular interaction theory. *Psychol Rev* **112**, 1000-21.
- Nagel, T. 1974 What is it like to be a bat? *Philosophical Review* **83**, 435-50.
- Nisbett, R. E. & Wilson, T. D. 1977 Telling more than we know: verbal reports on mental processes. *Psychological Review* **LXXXIV**, 231-259.
- Pally, R. 2005 Non-conscious prediction and a role for consciousness in correcting prediction errors. *Cortex* **41**, 643-62.
- Pisella, L., Grea, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., Boisson, D. & Rossetti, Y. 2000 An 'automatic pilot' for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. *Nat Neurosci* **3**, 729-36.
- Polger, T. in press Rethinking the evolution of consciousness. In *Blackwell companion to consciousness* (ed. S. Schneider & M. Velmans). Oxford: Blackwell.
- Polger, T. & Flanagan, O. 1995 Zombies and the function of consciousness. *Journal of Consciousness Studies* **2**, 313-321.

- Putnam, H. 1988 *Representation and reality*. Cambridge, MA: MIT Press.
- Revonsuo, A. 2005 *Inner Presence: Consciousness as a Biological Phenomenon*. Cambridge, MA: MIT Press.
- Robinson, W. S. 2007 Epiphenomenalism. In *The Stanford Encyclopedia of Philosophy (Spring 2007 Edition)* (ed. E. N. Zalta).
- Roger, M. & Galand, G. 1981 Operant conditioning of visual evoked potentials in man. *Psychophysiology* **18**, 477-82.
- Rosenthal, D. 2008 Consciousness and its function. *Neuropsychologia* **46**.
- Rosenthal, D. M. 2005 *Consciousness and mind*. Oxford: Clarendon.
- Sanz, R., Lopez, I., Rodriguez, M. & Hernandez, C. 2007 Principles for consciousness in integrated cognitive control. *Neural Netw* **20**, 938-46.
- Schneider, W., Pimm-Smith, M. & Worden, M. 1994 Neurobiology of attention and automaticity. *Curr Opin Neurobiol* **4**, 177-82.
- Schneider, W. & Shiffrin, R. M. 1977 Controlled and automatic human information processing I: Detection, search, and attention. *Psychological Review* **84**, 1-66.
- Searle, J. 1992 *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Seth, A. K. 2008 Causal networks in simulated neural systems. *Cognitive Neurodynamics* **2**, 49-64.
- Seth, A. K. & Baars, B. J. 2005 Neural Darwinism and consciousness. *Consciousness and Cognition* **14**, 140-168.
- Seth, A. K., Baars, B. J. & Edelman, D. B. 2005 Criteria for consciousness in humans and other mammals. *Consciousness and Cognition* **14**, 119-139.
- Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M. & Pessoa, L. 2008 Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends Cogn Sci* **12**, 314-21.
- Seth, A. K. & Edelman, G. M. 2007 Distinguishing causal interactions in neural populations. *Neural Comput* **19**, 910-933.
- Seth, A. K., Izhikevich, E., Reeke, G. N. & Edelman, G. M. 2006 Theories and measures of consciousness: An extended framework. *Proc Natl Acad Sci U S A* **103**, 10799-10804.
- Shanahan, M. 2006 A cognitive architecture that combines internal simulation with a global workspace. *Conscious Cogn* **15**, 433-49.
- Shanahan, M. & Baars, B. 2005 Applying global workspace theory to the frame problem. *Cognition* **98**, 157-76.
- Sherrington, C. S. 1906 *The integrative action of the nervous system*. New Haven, CT: Yale University Press.
- Sloman, A. & Chrisley, R. 2003 Virtual machines and consciousness. *Journal of Consciousness Studies* **10**, 133-72.
- Sporns, O., Tononi, G. & Edelman, G. M. 2000 Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cerebral Cortex* **10**, 127-141.
- Stadler, M. A. & Frensch, P. A. (ed.) 1997 *Handbook of implicit learning*: Sage.
- Tononi, G. 2004 An information integration theory of consciousness. *BMC Neurosci* **5**, 42.
- Tononi, G. & Edelman, G. M. 1998 Consciousness and complexity. *Science* **282**, 1846-51.

- Tononi, G., Sporns, O. & Edelman, G. M. 1994 A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci U S A* **91**, 5033-7.
- Trevethan, C. T., Sahraie, A. & Weiskrantz, L. 2007 Can blindsight be superior to 'sighted-sight'? *Cognition* **103**, 491-501.
- van Gulick, R. 2004 Consciousness. In *The Stanford Encyclopedia of Philosophy (Summer 2004 edition)* (ed. E. N. Zalta).
- Wegner, D. 2002 *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M. 2003 The mind's best trick: how we experience conscious will. *Trends Cogn Sci* **7**, 65-69.
- Weiskrantz, L. 1986 *Blindsight: A case study and implications*. Oxford: Clarendon Press.
- Wright, L. 1973 Functions. In *Philosophical Review*, vol. 82.
- Zeman, A. 2005 What in the world is consciousness. *Progress in Brain Research* **150**, 1-10.