



Published in final edited form as:

*Cogn Neurosci.* 2011 January 1; 2(2): 98–113. doi:10.1080/17588928.2011.565121.

## Human consciousness and its relationship to social neuroscience: A novel hypothesis

**Michael S. A. Graziano and Sabine Kastner**

Princeton University, Department of Psychology, Green Hall, Princeton NJ 08544

### Abstract

A common modern view of consciousness is that it is an emergent property of the brain, perhaps caused by neuronal complexity, and perhaps with no adaptive value. Exactly what emerges, how it emerges, and from what specific neuronal process, is in debate. One possible explanation of consciousness, proposed here, is that it is a construct of the social perceptual machinery. Humans have specialized neuronal machinery that allows us to be socially intelligent. The primary role for this machinery is to construct models of other people's minds thereby gaining some ability to predict the behavior of other individuals. In the present hypothesis, awareness is a perceptual reconstruction of attentional state; and the machinery that computes information about other people's awareness is the same machinery that computes information about our own awareness. The present article brings together a variety of lines of evidence including experiments on the neural basis of social perception, on hemispatial neglect, on the out-of-body experience, on mirror neurons, and on the mechanisms of decision-making, to explore the possibility that awareness is a construct of the social machinery in the brain.

---

“Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant...”

Hippocrates, Fifth Century, BC.

### Introduction

A common neuroscientific assumption about human consciousness is that it is an emergent property of information processing in the brain. Information is passed through neuronal networks, and by an unknown process consciousness of that information ensues. In such a view, a distinction is drawn between the information represented in the brain, that can be studied physiologically, and the as-yet unexplained property of being conscious of that information. In the present article a novel hypothesis is proposed that differs from these common intuitive notions. The hypothesis is summarized in the following five points.

First, when a person asserts “I am conscious of X,” whatever X may be, whether a color, a tactile sensation, a thought, or an emotion, the assertion depends on some system in the brain that must have computed the information, otherwise the information would be unavailable for report. Not only the information represented by X, visual information or auditory information for example, but also the essence of consciousness itself, the inner feeling attached to X, must be information or we would be unable to say that we have it. In this hypothesis, consciousness is not an emergent property, or a metaphysical emanation, but

is itself information computed by an expert system. This first point raises the question of why the brain would contain an expert system that computes consciousness. The question is addressed in the following points.

Second, people routinely compute the state of awareness of other people. A fundamental part of social intelligence is the ability to compute information of the type, “Bill is aware of X.” In the present proposal, the awareness we attribute to another person is our reconstruction of that person’s attention. This social capability to reconstruct other people’s attentional state is probably dependant on a specific network of brain areas that evolved to process social information, though the exact neural instantiation of social intelligence is still in debate.

Third, in the present hypothesis, the same machinery that computes socially relevant information of the type, “Bill is aware of X,” also computes information of the type, “I am aware of X.” When we introspect about our own awareness, or make decisions about the presence or absence of our own awareness of this or that item, we rely on the same circuitry whose expertise is to compute information about other people’s awareness.

Fourth, awareness is best described as a perceptual model. It is not merely a cognitive or semantic proposition about ourselves that we can verbalize. Instead it is a rich informational model that includes, among other computed properties, a spatial structure. A commonly overlooked or entirely ignored component of social perception is spatial localization. Social perception is not merely about constructing a model of the thoughts and emotions of another person, but also about binding those mental attributes to a location. We do not merely reconstruct that Bill believes this, feels that, and is aware of the other, but we perceive those mental attributes as localized within and emanating from Bill. In the present hypothesis, through the use of the social perceptual machinery, we assign the property of awareness to a location within ourselves.

Fifth, because we have more complete and more continuous data on ourselves, the perceptual model of our own awareness is more detailed and closer to detection threshold than our perceptual models of other people’s awareness.

The purpose of the present article is to elaborate on the hypothesis summarized above and to review some existing evidence that is consistent with the hypothesis. None of the evidence discussed in this article is conclusive. Arguably, little conclusive evidence yet exists in the study of consciousness. Yet the evidence suggests some plausibility to the present hypothesis that consciousness is a perception and that the perceptual model is constructed by social circuitry.

The article is organized in the following manner. First the hypothesis is outlined in greater detail (**Awareness as a product of social perception**). Second, a summary of recent work on the neuronal basis of social perception is provided (**Machinery for social perception and cognition**). A series of sections then describes results from a variety of areas of study, including hemispatial neglect, cortical attentional processing, aspects of self perception including the out-of-body illusion, mirror neurons as a possible mechanism of social perception, and decision-making as a means of answering questions about one’s own awareness. In each case the evidence is interpreted in light of the present hypothesis. One possible advantage of the present hypothesis is that it may provide a general theoretical basis on which to understand and fit together a great range of otherwise disparate and incompatible data sets.

## Awareness as a product of social perception

The hypothesis that consciousness is closely related to social ability has been suggested previously in many forms (e.g. Baumeister & Masicampo, 2010; Carruthers, 2009; Frith, 1995; Gazzaniga, 1970; Humphrey, 1983; Nisbett & Wilson, 1977). Humans have neuronal machinery that apparently contributes to constructing models of other people's minds (e.g. Brunet et al., 2000; Ciaramidaro et al., 2007; Gallagher et al., 2000; Samson et al., 2004; Saxe & Kanwisher, 2003; Saxe & Wexler, 2005). This circuitry may also contribute to building a model of one's own mind (e.g. Frith, 2002; Ochsner et al., 2004; Saxe et al., 2006; Vogeley et al., 2001; Vogeley et al., 2004). The ability to compute explicit, reportable information about our own emotions, thoughts, goals, and beliefs by applying the machinery of social cognition to ourselves can potentially explain self knowledge.

It has been pointed out, however, that self knowledge does not easily explain consciousness (Crick & Koch, 1990). Granted that we have self knowledge, and that we construct a narrative to explain our own behavior, how exactly do we become conscious of that information, and how does consciousness extend to other information domains such as colors, sounds, and tactile sensations? Constructing models of one's own mental processes could be categorized as "access consciousness" as opposed to "phenomenal consciousness" (Block, 1996). It could be considered a part of the "easy problem" of consciousness, determining the information of which we are aware, rather than the "hard problem" of determining *how* we become aware of it (Chalmers, 1995).

Social approaches to consciousness are not alone in these difficulties. Other theories of consciousness suffer from similar limitations. One major area of thought on consciousness focuses on the massive, brain-wide integration of information. For example, in his Global Workspace theory, Baars was one of the first to posit a unified, brain-wide pool of information that forms the contents of consciousness (Baars, 1983; Newman & Baars, 1993). A possible mechanism for binding information across brain regions, through the synchronized activity of neurons, was proposed by Singer and colleagues (Engel et al., 1990; Engel & Singer, 2001). Shortly after the first report from Singer and colleagues, Crick and Koch (1990) suggested that when information is bound together across regions of the cortex through the synchronized activity of neurons, it enters consciousness. Many others have since proposed theories of consciousness that include or elaborate on the basic hypothesis that consciousness depends on the binding of information (e.g. Grossbergm 1999; Lamme, 2006; Tononi, 2008; Tononi & Edelman, 1998). All of these approaches recognize that the content of consciousness includes a great complexity of interlinked information. But none of the approaches explain how it is that we become aware of that information. What exactly is the inner essence, the feeling of consciousness, that seems to be attached to the information?

Here we propose that the machinery for social perception provides that feeling of consciousness (Graziano, 2010). The proposal does not necessarily contradict previous accounts. It could be viewed as a way of linking social theories of consciousness with theories in which consciousness depends on informational binding. If consciousness is associated with a global workspace, or a bound set of information that spans many cortical areas, as so many others have suggested, then in the present proposal the awareness ingredient added to that global information set is provided by the machinery for social perception. In specific, awareness is proposed to be a rich descriptive model of the process of attention.

The proposal begins with the relationship between awareness and attention. The distinction between awareness and attention has been studied before (e.g. Dehaene et al., 2006; Jiang et

al., 2006; Kentridge et al., 2004; Koch & Tsuchiya, 2007; Lamme, 2004; Naccache et al., 2002). The two almost always covary, but under some circumstances it is possible to attend to a stimulus and at the same time be unaware of the stimulus (Jiang et al., 2006; Kentridge et al., 2004; Naccache et al., 2002). Awareness, therefore, is not the same thing as attention, but puzzlingly the two seem redundant much of the time. Here we propose an explanation for the puzzling relationship between the two: awareness is a perceptual model of attention. Like most informational models in the brain, it is not a literal transcription of the thing it represents. It is a caricature. It exaggerates useful, need-to-know information. Its purpose is not to provide the brain with a scientifically accurate account of attention, but to provide useful information that can help guide behavior. In the following paragraphs we discuss first the social perception of someone else's attentional state, and then the perception of one's own attentional state.

Arguably one of the most basic tasks in social perception is to perceive the focus of somebody else's attention. The behavior of an individual is driven mainly by the items currently in that individual's focus of attention. Hence computing that someone is attending to this visual stimulus, that sound, this idea, and that emotion, provides critical information for behavioral prediction. The importance of computing someone else's state of attention has been emphasized by others, and forms the basis for a body of work on what is sometimes called "social attention" (e.g. Birmingham & Kingstone, 2009; Friesen & Kingstone, 1998; Frischen et al., 2007; Nummenmaa & Calder, 2008; Samson et al., 2010). One of the visual cues used to perceive someone else's attentional state is the direction of gaze. Neurons that represent the direction of someone else's gaze have been reported in cortical regions thought to contribute to social perception including, in particular, area STS of monkeys and humans (Calder et al., 2002; Hoffman & Haxby, 2000; Perrett et al., 1985; Puce et al., 1998; Wicker et al., 1998). Gaze is of course not the only cue. A variety of other cues such as facial expression, body posture, and vocalization, presumably also contribute to perceiving the focus of somebody else's attention.

In the hypothesis proposed here, when we construct a perceptual model of someone else's focus of attention, that informational model describes awareness originating in that person and directed at a particular item. In this hypothesis, the brain explicitly computes an awareness construct, and awareness is the perceptual reconstruction of attention. Figure 1 provides an example to better explain this proposed relationship between awareness and attention.

In Figure 1, Abel looks at Bill and Bill looks at the coffee cup. First consider Bill, whose visual attention is focused on the cup. It is now possible to provide a fairly detailed account of visual attention, which has been described as a process by which one stimulus representation wins a neuronal competition among other representations (for review, see Beck & Kastner, 2009; Desimone & Duncan, 1995). The competition can be influenced by a variety of signals. For example, bottom up signals, such as the brightness or the sudden onset of a stimulus, may cause its representation to win the competition and gain signal strength at the expense of other representations. Likewise, top-down signals that emphasize regions of space or that emphasize certain shapes or colors may be able to bias the competition in favor of one or another stimulus representation. Once a stimulus representation has won the competition, and its signal strength is boosted, that stimulus is more likely to drive the behavior of the animal. This self-organizing process is constantly shifting as one or another representation temporarily wins the competition. In Figure 1, Bill's visual system builds a perceptual model of the coffee cup that wins the attentional competition.

Now consider Abel, whose machinery for social perception constructs a model of Bill's mind. This model includes, among other properties, the following three pieces of information. First, awareness is present. Second, the awareness emanates from Bill. Third, the awareness is directed in a spatially specific manner at the location of the cup. These properties — the property of awareness and the two spatial locations to which it is referred — are perceptual constructs in Abel's brain.

In this formulation, Bill's visual attention is an event to be perceived, and awareness is the perceptual counterpart to it constructed by Abel's social machinery. Note the distinction between the reality (Bill's attentional process) and the perceptual representation of the reality (Abel's perception that Bill is aware). The reality is quite complex. It includes the physics of light entering the eye, the body orientation and gaze direction of Bill, and a large set of unseen neuronal processes in Bill's brain. The perceptual representation of that reality is much simpler, containing an amorphous somewhat ethereal property of awareness that can be spatially localized at least vaguely to Bill and that, in violation of the physics of optics, emanates from Bill toward the object of his awareness. (For a discussion of the widespread human perception that vision involves something coming out of the eyes, see Cottrell & Winer, 1994; Gross, 1999.) The perceptual model is simple, easy, implausible from the point of view of physics, but useful for keeping track of Bill's state and therefore for helping to predict Bill's behavior. As in all perception, the perception of awareness is useful rather than accurate.

Consider now the modified situation in which Abel and Bill are the same person. A person is never outside of a social context because he is always with himself and can always use his considerable social machinery to perceive, analyze, and answer questions about himself. Abel/Bill focuses visual attention on the coffee cup. Abel/Bill also constructs a model of the attentional process. The model includes the information: awareness is present; the awareness emanates from me; the awareness is directed at the cup. If asked, "Are you aware of the cup?" Abel/Bill can cognitively scan the contents of this model and on that basis answer, "Yes."

If asked, "What exactly do you mean by awareness of the cup?" Abel/Bill can again scan the informational model, abstract properties from it, and report something like, "My awareness is a feeling, a vividness, a mental seizing of the stimulus. My awareness feels like it is located inside me. In a sense it *is* me. It is my mind apprehending something." These summaries reflect the brain's model of the process of attention.

Awareness, in this account, is one's social intelligence perceiving one's focus of attention. It is a second-order representation of attention. In that sense the hypothesis may seem similar to proposals involving metacognition (e.g. Carruthers, 2009; Pasquali et al., 2010; Rosenthal, 2000). Metacognition generally refers to semantic knowledge about one's mental processes, or so-called "thinking about thinking". The proposal here, however, is different. While people clearly have semantic knowledge about their attentional state, what is proposed here is specifically the presence of a rich, descriptive, perceptual model of attentional state that, like most perception, is computed involuntarily and is continuously updated. When we gain cognitive access to that perceptual model and summarize it in words, we report it as awareness. Block (1996) distinguished between phenomenal consciousness (the property of consciousness itself) and access consciousness (cognitive access to the property of consciousness). In the present theory, the perceptual representation of attentional state is akin to phenomenal consciousness. The cognitive access to that representation, that allows us abstract semantic knowledge and to report on it, is akin to access consciousness.

In the present hypothesis we propose a similarity between perceiving someone else's awareness and perceiving one's own awareness. Both are proposed to be social perceptions dependant on the same neuronal mechanisms. Yet do we really perceive someone else's awareness in the same sense that we perceive our own, or do we merely acknowledge in an abstract or cognitive sense that the other person is likely to be aware? In a face-to-face conversation with another person, so many perceptions and cognitive models are present regarding tone of voice, facial expression, gesture, and the semantic meaning of the other person's words, that it is difficult to isolate the specific perceptual experience of the other person's awareness. Yet there is one circumstance in which extraneous perceptions are minimized and the perception of someone else's awareness is relatively isolated and therefore more obvious. This circumstance is illustrated in Figure 2. Everyone is familiar with the spooky sensation that someone is staring at you from behind (Coover, 1913; Titchner, 1898). Presumably built on lower level sensory cues such as subtle shadows or sounds, the perception of a mind that is located behind you and that is aware of you is a type of social perception and a particularly pure case of the perception of awareness. Other aspects of social perception are stripped away. The perceptual illusion includes a blend of three components: the perception that awareness is present, the perception that the awareness emanates from a place roughly localized behind you, and the perception that the awareness is directed at a specific object (you). This illusion helps to demonstrate that awareness is not only something that a brain perceives to be originating from itself — I am aware of this or that — but something that a brain can perceive as originating from another source. In the present argument, awareness is a perceptual property that can be attributed to someone else's mind or to one's own mind.

The present hypothesis emerges from the realization that social perception is not merely about reconstructing someone else's thoughts, beliefs, or emotions, but also about determining the state of someone else's attention. Information about someone else's attention is useful in predicting the likely moment-by-moment behavior of the person. The social machinery computes that Bill is aware of this, that, and the other. Therefore social perception, when applied to oneself, provides not only a description of one's own inner thoughts, beliefs, and feelings, but also a description of one's awareness of items in the outside environment. It is for this reason that awareness, awareness of anything, awareness of a color, or a sound, or a smell, not just self-awareness, can be understood as a social computation.

The examples given above focus on visual attention and visual awareness. The concept, however, is general. In the example in Figure 1, Bill could just as well attend to a coffee cup, a sound, a feeling, a thought, a movement intention, or many other cognitive, emotional, and sensory events. In the present hypothesis, awareness is the perceptual reconstruction of attention, and therefore anything that can be the subject of attention can also be the subject of awareness.

## **Machinery for social perception and cognition**

Arguably social neuroscience began with the discovery by Gross and colleagues of hand and face cells in the inferior temporal cortex of monkeys (Desimone et al., 1984; Gross et al., 1969). Further work indicated that a neighboring cortical area, the Superior Temporal Polysensory area (STP), contains a high percentage of neurons tuned to socially relevant visual stimuli including faces, biological motion of bodies and limbs, and gaze direction (Barraclough et al., 2006; Bruce et al., 1981; Jellema & Perrett, 2003; Jellema & Perrett, 2006; Perrett et al., 1985). In humans, through the use of functional magnetic resonance imaging (fMRI), a region that responds more strongly to the sight of faces than to other objects was identified in the fusiform gyrus (Kanwisher et al., 1997). Areas in the human

superior temporal sulcus (STS) were found to become active during the perception of gaze direction and of biological motion such as facial movements and reaching (Grossman et al., 2000; Pelphrey et al., 2005; Puce et al., 1998; Thompson et al., 2007; Vaina et al., 2001; Wicker et al., 1988). This evidence from monkeys and humans suggests that the primate visual system contains a cluster of cortical areas that specializes in processing the sensory cues related to faces and gestures that are relevant to social intelligence.

Other studies in social neuroscience have investigated a more cognitive aspect to social intelligence sometimes termed theory-of-mind (Frith & Frith, 2003; Premack & Woodruff, 1978). Tasks that require the construction of models of the contents of other people's minds have been reported to engage a range of cortical areas including the STS, the temporo-parietal junction (TPJ), and the medial prefrontal cortex (MPFC), with a greater but not exclusive activation in the right hemisphere (Brunet et al., 2000; Ciaramidaro et al., 2007; Fletcher et al., 1995; Gallagher et al., 2000; Goel et al., 1995; Saxe & Kanwisher, 2003; Saxe & Wexler, 2005; Vogley et al., 2001).

The relative roles of these areas in social perception and cognition are still in debate. It has been suggested that the TPJ is selectively recruited during theory-of-mind tasks, especially during tasks that require constructing a model of someone else's beliefs (Saxe & Kanwisher, 2003; Saxe & Wexler, 2005). Damage to the TPJ is associated with impairment in theory-of-mind reasoning (Apperly et al., 2004; Samson et al., 2004; Weed et al., 2010).

The STS, adjacent to the TPJ, has been argued to play a role in perceiving someone else's movement intentions (Blakemore et al., 2003; Pelphrey et al., 2004; Wyk et al., 2009). Not only is the STS active during the passive viewing of biological motion, as noted above, but the activity distinguishes between goal-directed actions such as reaching to grasp an object and non-goal-directed actions such as arm movements that do not terminate in a grasp (Pelphrey et al., 2004). Even when a subject views simple geometric shapes that move on a computer screen, movements that are perceived as intentional activate the STS whereas movements that appear mechanical do not (Blakemore et al., 2003).

The role of the MPFC is not yet clear. It is consistently recruited in social perception tasks and theory-of-mind tasks (Brunet et al., 2000; Fletcher et al., 1995; Frith, 2002; Gallagher et al., 2000; Goel et al., 1995; Passingham et al., 2010; Vogley et al., 2001) but lesions to it do not cause a clear deficit in theory-of-mind reasoning (Bird et al., 2004). Some speculations about the role of the MPFC in social cognition are discussed in subsequent sections.

Taken together these studies suggest that a network of cortical areas, mainly but not exclusively in the right hemisphere, collectively build models of other minds. Different areas within this cluster may emphasize different aspects of the model, though it seems likely that the areas interact in a cooperative fashion.

The view that social perception and cognition are emphasized in a set of cortical areas dedicated to social processing is not universally accepted. At least two main rival views exist. One view is that the right TPJ and STS play a more general role related to attentional processing rather than a specific role related to social cognition (e.g. Astafiev et al., 2006; Corbetta et al., 2000; Mitchell, 2008; Shulman et al., 2010). A second alternative view is that social perception is mediated at least partly by mirror neurons in the motor system that compute one's own actions and also simulate the observed actions of others (e.g. Rizzolatti & Sinigaglia, 2010). Both of these alternative views are discussed in subsequent sections. Much of the discussion below, however, is based on the hypothesized role of the right TPJ and STS in social perception and social cognition.

## **Prediction 1: Damage to the machinery for social perception should cause a deficit in awareness**

If the present proposal is correct, if awareness is a construct of the machinery for social perception, then damage to the right TPJ and STS, the brain areas most associated with constructing perceptual models of other minds, should sometimes cause a deficit in one's own awareness. These cortical regions are heterogeneous. Even assuming their role in social cognition, different sub areas probably emphasize different functions. As discussed in the next section, the sub regions of TPJ involved in attention may be partially distinct from sub regions involved in theory-of-mind (Scholz et al., 2009). Therefore, even in the present hypothesis, a lesion to the TPJ and STS should not always impact all aspects of social cognition equally. A range of symptoms might result. The present hypothesis does, however, make a clear prediction: damage to the right TPJ and STS should often be associated with a deficit in consciousness.

The clinical syndrome that comes closest to an awareness deficit is hemispatial neglect, the loss of processing of stimuli usually on the left side of space after damage to the right hemisphere of the brain (Brain, 1941; Critchley, 1953). Patients classically fail to report, react to, or notice anything on the left half of space, whether visual, auditory, tactile, or memory. The left half of space, and any concept that it ever existed, are erased from the patient's awareness.

It is now generally accepted that there is no single neglect syndrome. A range of lesion sites can result in neglect and different neglect patients can have somewhat different mixtures of symptoms (Halligan & Marshall, 1992; Halligan et al., 2003; Vallar, 2001). It is therefore not correct to attribute neglect to a single brain area or mechanism. It has been reported, however, that a strong form of neglect, the almost total loss of conscious acknowledgement of the left side of space or anything in it, occurs most often after damage to the right TPJ (Valler & Perani, 1986). In at least one subset of neglect patients who lacked any accompanying low-level blindness, and in this sense were more "pure" in their neglect symptoms, the most common lesion site was ventral and anterior to the TPJ, in the right STS (Karnath et al., 2001). Temporary interference with the right TPJ using transcranial magnetic stimulation (TMS) has also been reported to induce symptoms of left hemispatial neglect (Meister et al., 2006).

Some have argued that neglect is more commonly associated with damage to the posterior parietal lobe rather than to the more ventral TPJ or STS (Mort et al., 2003). A parietal locus for neglect is certainly a more traditional view (Brain, 1941; Critchley, 1953; Gross & Graziano, 1995). Neglect symptoms can also be observed after frontal lesions, though they tend to be less severe (Heilman & Valenstein, 1972a; Mesulam, 1999; Ptak & Schnider, 2010). These parietal and frontal sites for neglect are consistent with the proposal of a parieto-frontal network for the top-down control of attention (Ptak & Schnider, 2010; Szczepanski et al., 2010). How can parieto-frontal sites for neglect be reconciled with the observation of severe neglect from lesions in the TPJ and STS? One of the primary reasons for these differences among studies may be a disagreement over the definition of neglect. Different tests for neglect may result in a selection of different patient populations and therefore different observed lesion sites (Ferber & Karnath, 2001; Rorden et al., 2006). In particular, the two most common clinical tests for neglect may measure different deficits. In a line cancellation test, the patient crosses out short line segments scattered over a visual display. This test measures awareness because the patient cancels only the line segments that reach awareness and fails to cancel the line segments that do not reach awareness. In contrast, in a line bisection task, the patient attempts to mark the center of a long horizontal line. This test does not measure awareness or the lack thereof, since the patient is always



aware of the horizontal line. Instead the test measures a relative response bias toward one side, perhaps caused by an underlying attentional bias. Different variants of neglect, revealed by these different tests, might be associated with different lesion sites. It is of interest, given the present hypothesis, that neglect defined as an awareness deficit through the use of line cancellation was associated with more ventral lesions in the STS (Ferber & Karnath, 2001; Karnath et al., 2001; Rorden et al., 2006).

The point of the present section is not that neglect is “really” an awareness deficit instead of an attentional bias, and not that it is “really” caused by lesions to the right TPJ or STS instead of to the right inferior parietal lobe. Undoubtedly different symptoms and lesions sites can fall under the more general label of hemispatial neglect. The present hypothesis predicts that there should be at least two different kinds of neglect associated with two different brain systems, one associated with the process of *controlling* attention (neglect caused by damage to parieto-frontal attentional mechanisms) and the other associated with the process of perceptually *representing* attention (neglect caused by TPJ and STS damage).

Does the unilateral nature of neglect argue against the present interpretation? Shouldn't a consciousness area, if damaged, lead to a total loss of conscious experience and not merely a unilateral loss? We do not believe the laterality of neglect argues against the present interpretation. If social perception and social cognition were found to activate a right hemisphere region only, then damage to that region, according to the present theory, might eliminate all consciousness. The studies reviewed above on the neuronal basis of social perception and social cognition, however, suggest that these functions are represented in a bilateral manner with a strong emphasis on the right side. If consciousness is a construct of the social machinery, as suggested here, then it is portioned in some unequal manner between the hemispheres. One would expect, therefore, that damage to the system for social perception would, depending on the hemisphere, have an asymmetric effect on consciousness.

Some controversy surrounds the exact explanation for the unilateral nature of neglect. One view is a representational hypothesis (Heilman & Valenstein, 1972b; Mesulam, 1981; Mesulam, 1999). In that hypothesis, the right hemisphere contains some critical type of representation needed for awareness, that covers both sides of space, whereas the left hemisphere contains a representation only of the right side of space. Damage to the right hemisphere therefore leaves the patient without awareness of the left side of space, whereas damage to the left hemisphere leaves the patient mainly behaviorally intact. An alternative explanation is based on the concept of inter-hemispheric competition among controllers of attention (Kinsbourne, 1977; Szczepanski et al., 2010). In that view, damage to the parietal lobe on one side reduces the ability of the attentional control system that serves the opposite side of space. An imbalance is created and attention is unavoidably drawn to the ipsilesional side of space. The present theory of consciousness may be able to accommodate both of these seemingly conflicting views of the mechanisms of neglect. In the present theory, the brain contains at least two general processes related to attention. One is the control of attention, perhaps more emphasized in a parietal-frontal network, lesions to which may result in a more competition-style imbalance of attention; and the other is the perceptual representation of attention, emphasized in more ventral areas involved in social perception including TPJ and STS, damage to which may result in a more representational-style neglect.

## **Prediction 2: The machinery for social perception should carry signals that correlate with attention**

In the present hypothesis, a basic task of social perception is to reconstruct the focus of someone else's attention. In the same manner, in perceiving oneself, the social machinery reconstructs one's own constantly changing focus of attention. A prediction that follows from this hypothesis is that tasks that involve focusing or shifting attention should evoke brain activity not only in areas that participate in attentional control (such as the parieto-frontal attention network; for review see Beck & Kastner, 2009), but also in the social circuitry that generates a reconstruction of one's focus of attention. The prediction is therefore that the right TPJ and STS should be active in association with changes in attention.

Area TPJ and the adjacent, caudal STS show elevated activity in response to stimuli that are unexpected, that appear at unexpected locations, or that change in an unexpected manner (Astafiev et al., 2006; Corbetta et al., 2000; Shulman et al., 2010). This response to salient events is strongest in the right hemisphere. It has therefore been suggested that TPJ and STS are part of a right-lateralized system, a "ventral network" that is involved in some aspects of spatial attentional processing (Astafiev et al., 2006; Corbetta et al., 2000; Shulman et al., 2010).

The attention-related responses found in the STS and TPJ could be viewed as contradictory to the social functions proposed for the same areas. It could be argued that the attention findings challenge the view that right STS and TPJ are in any way specialized for social processing. Perhaps they serve a more general function related to attention. In the present hypothesis, however, the two proposed functions are compatible. There is no contradiction. The social machinery is constantly computing and updating a model of one's own mind, and central to that model is a representation of one's focus of attention. Tracking an individual's attention is a fundamental task in social perception.

An alternate explanation for why attentional functions and social functions are both represented in the right TPJ and STS is that the two functions are emphasized in different cortical areas that happen to be so near each other that fMRI cannot easily resolve them. One functional imaging study indicated that the two types of tasks recruited overlapping areas of activity in the TPJ with spatially separated peaks (Scholz et al., 2009). This attempt to separate the "attention" sub-area of TPJ from the "theory of mind" sub-area, however, does not entirely address the underlying issue of proximity. The cortex is generally organized by functional proximity. Similar or related functions tend to be processed near each other (Aflalo & Graziano, 2011; Graziano & Aflalo, 2007). This trend toward functional clustering is certainly true in the social domain. Even if some spatial separation exists between attentional functions and theory-of-mind functions, why are the two functions clustered so closely within the right TPJ and STS? Rather than view this proximity of different functions as a contradiction, or suggest that one function must be correct and the other a mistake, or suggest that the functions must be separate from each other and co-localized merely as an accident of brain organization, it is suggested here that the functions share a common cause and therefore are not mutually contradictory. In the present hypothesis, reconstructing someone's beliefs (such as in a false belief task) is only one component of social perception. Another component is tracking someone's attention; a third is constructing a model of someone's movement intentions. Presumably many other components of social perception exist, and these components may be represented cortically in adjacent, partially separable regions.

### **Prediction 3: The machinery for social perception should construct a self model**

A direct prediction of the present hypothesis is that the cortical machinery that constructs models of other minds should also construct a perceptual model of one's own mind.

Some of the most compelling evidence on the self model involves the out-of-body illusion, in which the perceived location of the mind no longer matches the actual location of the body. It is essentially an error in constructing a perceptual model of one's own mind. Perceptual mislocalizations of the self can be induced in normal, healthy people by cleverly manipulating visual and somatosensory information in a virtual-reality setup (Ehrsson, 2007; Lenggenhager et al., 2007). A complete out-of-body illusion can also be induced with electrical stimulation of the cortex. In one experiment, Blanke et al. (2002) electrically stimulated the cortical surface in an awake human subject. When electrical stimulation was applied to the TPJ in the right hemisphere, an out-of-body experience was induced. The stimulation evidently interfered with the machinery that normally assigns a location to the self. Subsequent experiments showed that tasks that involve mentally manipulating one's spatial perspective evoke activity in the right TPJ (Blanke et al., 2005; Zacks et al., 2003) and are disrupted by transcranial magnetic stimulation to the right TPJ (Blanke et al., 2005). The evidence suggests that a specific mechanism in the brain is responsible for building a spatial model of one's own mind, and that the right TPJ plays a central role in the process. Given that the right TPJ has been implicated so strongly in the social perception of others, this evidence appears to support the hypothesis that the machinery for social perception also builds a perceptual model of one's own mental experience.

An entire subfield of neuroscience and psychology is devoted to the subject of the body schema (for review, see Graziano & Botvinik, 2002). For example, experiments have examined the neuronal mechanisms by which we know the locations of our limbs, incorporate tools into the body schema, or experience body shape illusions. However, among the many studies on the body schema, the particular phenomenon of the out-of-body experience is uniquely relevant to the discussion of awareness. The reason is that the out-of-body experience is specifically a spatial mis-localization of the source of one's awareness.

The very existence of an out-of-body illusion could be interpreted as evidence in support of the present hypothesis, for the following reasons. Suppose that awareness is not explicitly computed, as hypothesized here, but instead is an emergent property produced as a byproduct of neuronal activity. Neurons compute and transmit information, and as a result, somehow awareness of that information occurs. Why would we feel the awareness as located in, and emanating from, our own selves? By what mechanism does the awareness feel as though it is anchored here or anywhere else? The fact that awareness comes with a perceived location suggests that it is a perceptual model that, like all perception, is associated with a computed spatial location. Usually one does not notice that awareness has a perceived location. Its spatial structure, emanating from inside one's own body, is so obvious and so ordinary that we take it for granted. The out-of-body illusion, in which the spatial computation has gone awry, allows one to realize that awareness comes with a computed spatial arrangement. Evidently perceptual machinery in the brain computes one's awareness, assigns it a perceived source inside one's body, and interference with the relevant circuitry results in an error in the computation. The out-of-body experience demonstrates that awareness is a computation subject to error.

In a second general approach to the self model, in a series of experiments subjects performed self perception tasks such as answering questions about their own feelings or beliefs or monitoring their own behavior while brain activity was measured through

functional imaging (Frith, 2002; Ochsner et al., 2004; Passingham et al., 2010; Saxe et al., 2006; Vogeley et al., 2001). The network of areas recruited during self perception included the right TPJ, STS, and MPFC. The activity in the MPFC was the largest and most consistent. Again, the findings appear to support the general hypothesis that the machinery for social perception is recruited to build a perceptual model of the self. These functional imaging studies on the self model, however, are less pertinent to the immediate question of awareness. The reason is that the studies examined self-report of emotions or thoughts, not the report of awareness itself. The focus of the present proposal is not self-awareness in specific, but awareness in general. It is extremely difficult to pinpoint the source of awareness in a functional imaging study of this type, because regardless of the task, whether one is asked to be aware of one's own mind state or of someone else's, one is nonetheless aware of something, and therefore the computation of awareness is present in both experimental and control tasks.

### Challenge 1: How can the inner “feeling” of awareness be explained?

Thus far we have presented a novel hypothesis about human consciousness, outlined three general predictions that follow from the hypothesis, and reviewed findings that are consistent with those predictions. The evidence is by no means conclusive, but does suggest that the hypothesis has some plausibility. Human consciousness may be a product of the same machinery that builds social perceptual models of other people's mental states. Arguably the strongest evidence in favor of the hypothesis is that damage to that social machinery causes a profound deficit in consciousness. The hypothesis, however, also has some potential challenges. Several aspects of conscious experience are not, at the outset, easily explainable by the present hypothesis. We address three of these potential challenges to the hypothesis. In each case we offer a speculation about how the hypothesis might explain, or at least be compatible with, the known phenomena. These sections are necessarily highly speculative, but the issues must be addressed for the present hypothesis to have any claim to plausibility.

The first challenge we take up concerns the difference between constructing an informational model about awareness and actually “feeling” aware. In the present hypothesis, networks that are expert at social computation analyze the behavior of other people and compute information of the type, “Bill is aware of X.” The same networks, by hypothesis, compute information of the type, “I am aware of X,” thereby allowing one to report on one's own awareness. Can such a machine actually “feel” aware or does it merely compute an answer without any inner experience?

The issue is murky. If you ask yourself, “Am I merely a machine programmed to answer questions about awareness, or do I actually *feel* my own awareness,” you are likely to answer yes, you actually feel it. You might even specify that you feel it inside you as a somewhat amorphous but nonetheless real thing. But in doing so, you are merely computing the answers to other questions. Distinguishing between actually having an inner experience, and merely computing, when asked, that you have an inner experience, is a difficult, perhaps impossible task.

In the present account, we are not hypothesizing that a set of semantic, symbolic, or linguistic propositions can take the place of the inner essence of awareness, but that a representation, an informational picture, comprises awareness. In the distinction between phenomenal consciousness and access consciousness suggested by Block (1996), the informational representation proposed here is similar to phenomenal consciousness, and our ability to cognitively access that representation and answer questions about it is similar to access consciousness.

Yet even assuming a rich representational model of attention, and even assuming that the model comprises what we report to be awareness, why does it “feel” like something to us? Why the similarity between awareness and feeling? To address the issue we begin with a set of illusions, both related to the spatial structure assigned to the model of awareness.

As mentioned in the previous section, the out-of-body experience is the illusion of floating outside your own body. The perceived source of your awareness no longer matches the actual location of your body. The illusion can be induced by direct electrical stimulation applied to the right TPJ (Blanke et al., 2002). Similar mislocalizations of the self can be induced in people by manipulating visual and somatosensory information in a virtual-reality setup (Ehrsson, 2007; Lenggenhager et al., 2007). Evidently a specific mechanism in the brain is responsible for embodiment, or building a spatial model of one’s own awareness.

Does a similar property of embodiment apply when constructing a model of someone else’s awareness? The illusion illustrated in Figure 2 suggests that a model of someone else’s awareness does come with an assigned location. The perception that someone is staring at you from behind (Coover, 1913; Titchner, 1898) includes a spatial structure in which awareness emanates from a place roughly localized behind you and is directed at you.

These illusions demonstrate a fundamental point about social intelligence that is often ignored. When building a model of a mind, whether one’s own or someone else’s, the process is not merely one of inferring disconnected mental attributes — beliefs, emotions, intentions, awareness. The model also has a spatial embodiment. The model, like any other perceptual model, consists of a set of attributes assigned to a location. Social perception is not normally compared directly to sensory perception, but in this manner they are similar. They both involve computed properties — mental properties or sensory properties — as well as a computed location, bound together to form a model of an object.

In this hypothesis, one’s own consciousness is a perceptual model in which computed properties are attributed to a location inside one’s body. In this sense consciousness is another example of somesthesia, albeit a highly specialized one. It is a perceptual representation of the workings of the inner environment. Like the perception of a stomachache, or the perception of joint rotation, or the perception of light-headedness, or the perception of cold and hot, the perception of one’s own awareness is a means of monitoring processes inside of the body.

A central philosophical question of consciousness could be put this way: Why does thinking *feel* like something? When Abel solves a math problem in his head, why doesn’t he merely process the numerical information and output the number without feeling it? Why does he feel as though a process is occurring inside his head? To rephrase the question more precisely: Why is it that, on introspection, when he engages a decision process to compare thinking to somatosensory processing, he consistently concludes that a similarity exists? We speculate that a part of the answer may be that his awareness of his own thinking is a model of the internal environment, assigned to a location inside his body, and as such is a type of body perception. He therefore reports that thinking shares associations with processing in the somatosensory domain.

Consider again the illusion diagrammed in Figure 2, the feeling that somebody is looking at you from behind. Subjectively, you feel something on the back of your neck or in your mind that seems to warn you. You feel the presence of awareness. But it is not your own awareness. It is a perceptual model of someone else’s awareness. How can the feeling be explained? The perceptual model includes a property (awareness) and two spatial locations (the source of the awareness behind you and the focus of the awareness on you). One of these spatial locations overlaps with your self-boundary. It “feels” like something in the

sense that a property is assigned to a location inside of you. It literally belongs to the category of somesthesis because it is a perceptual reconstruction of the internal environment. What is key here is not merely localizing the event to the inside of the body, but the rich informational associations that come with that localization, the associations with other forms of somesthesis including touch, temperature, pressure, and so on. In this speculation, the perceptual representation of someone else's attention on the back of one's head comes with a vast shadowy complex of information that is linked to it and is subtly activated along with it.

In addition to the circumstance of "feeling" someone else's awareness directed at you, consider the opposite circumstance in which your own awareness is directed at something else. Do you "feel" your own awareness of the item? In the present hypothesis, your machinery for social perception constructs a model of your own attentional processes. That self model includes the property of awareness, a source of the awareness inside you, and a focus of the awareness on the object. One of these spatial locations overlaps with your self boundary. On introspection you can decide that you "feel" the source of your own awareness, in the sense that it is a perceptual model of your internal environment and therefore shares rich informational associations with other forms of somesthesis.

In summary, we speculate that even though conscious experience is information computed by expert systems, much like the information in a calculator or a computer, it none-the-less can "feel" like something in at least the following specific sense: a decision process, accessing the self model, arrives at the conclusion and triggers the report that awareness shares associations with somesthesis. This account, of course, does not explain conscious experience in its entirety; but it helps to make a central point of the present perspective. 'Experienceness' itself may be a complex weaving together of information, and it is possible to tease apart at least some of that information (such as the similarity to somesthesis) and understand how it might be computed.

## **Challenge 2: How can the machinery for social perception gain access to modality-spanning information?**

Social perception presents a challenge to neuronal representation. In color perception, one can study the areas of the brain that receive color information. In auditory perception, one can study auditory cortex. But in social perception, a vast range of information must converge. Is the other person aware of that coffee cup? Is he aware of the chill in the room? Is he aware of the abstract idea that I am trying to get across, or is he distracted by his own thoughts? What emotion is he feeling? Social perception requires an extraordinary multimodal linkage of information. All things sensory, emotional, or cognitive, that might affect another person's behavior, must be brought together and considered, in order to build a useful predictive model of the other person's mind. Hence the neuronal machinery for social perception must be an information nexus. It is implausible to talk about a specific region of the brain that performs all of social perception. If the TPJ and STS play a central role, which seems likely on the evidence reviewed above, they must serve as nodes in a brain-wide network.

In the present hypothesis, awareness is part of a social perceptual model of one's own mind. Consider the case in which you report, "I am aware of the green apple." Which part of this linked bundle of information is encoded in the social machinery, perhaps including the TPJ or STS, and which is encoded elsewhere in the brain? Can TPJ or STS contain a complete unified representation of the awareness of green apples? Or, instead, does the social circuitry compute the "awareness" construct, the visual cortex represent the "green apple," while a binding process links the two neuronal representations? On sheer speculation, the second

possibility seems more likely. Certainly the importance to consciousness of binding information across wide regions of the brain has been discussed before (e.g. Baars, 1983; Crick & Koch, 1990; Damasio, 1990; Engel & Singer, 2001; Grossberg, 1999; Lamme, 2006; Llinas & Ribary, 1994; Newman & Baars, 1993; Treisman, 1988; von der Malsburg, 1997). From the perspective of information theory, it has been suggested that consciousness is massively linked information (Tononi, 2008; Tononi & Edelman, 1998). The contribution of the present hypothesis is to suggest that binding of information across cortical areas is not, by itself, the raw material of awareness, but instead awareness is specific information, a construct about the nature of experience, that is computed by and represented in specific circuitry and that can be bound to larger network-spanning representations.

In Treisman's feature integration theory (Treisman & Gelade, 1980), when we perceive an object, different informational features such as shape, color, and motion, are bound together to form a unified representation. That binding requires attention. Without attention to the object, binding of disparate information about the object is possible but incomplete and often in error. If awareness acts like a feature that can be bound to an object representation, then attention to an object should be necessary for consistent or robust awareness to be attached to the object. In contrast, awareness of an object should not be necessary for attention to the object. In other words, it should be possible to attend to an object without awareness of it, but difficult to be aware of an object without attention to it. This pattern broadly matches the literature on the relationship between attention and awareness (e.g. Dehaene et al., 2006; Jiang et al., 2006; Kentridge et al., 2004; Koch & Tsuchiya, 2007; Lamme, 2004; Naccache et al., 2002).

Koch & Tsuchiya (2007) argued that, qualitatively, it seems possible to be aware of stimuli at the periphery of attention, and therefore awareness must be possible with minimal attention. But note that in this circumstance, speaking qualitatively again, one tends to feel aware of something without knowing exactly what the something is. Consider the proverbial intuition that something's wrong, or something's present, or something is intruding into awareness at the edge of vision, without being able to put your finger on what that item is exactly. This feeling could be described as awareness that is not fully or reliably bound to a specific item. Only on re-directing attention does one become reliably aware of the item. In this sense awareness acts precisely like the unreliable or false conjunctions occurring outside the focus of awareness in feature integration theory. It acts like computed information about an object, like a feature that, without attention, is not reliably bound to the object representation.

In the present hypothesis, therefore, the relationship between attention and awareness is rather complex. Not only is awareness a perceptual reconstruction of attention, but to bind awareness to a stimulus representation requires attention, because attention participates in the mechanism of binding.

### **Challenge 3: Is the present hypothesis compatible with simulation theory?**

In the present hypothesis, the human brain evolved mechanisms for social perception, a type of perception that allows for predictive modeling of the behavior of complex, brain-controlled agents. There is no assumption here about whether perception of others or perception of oneself emerged first. Presumably they evolved at the same time. Whether social perception is applied to oneself or to someone else, it serves the adaptive function of a prediction engine for human behavior.

An alternative view is that awareness emerged in the human brain for reasons unknown, perhaps as an epiphenomenon of brain complexity. Social perception was then possible by the use of empathy. We understand other people's minds by comparison to our own inner

experience. This second view has the disadvantage that it does not provide any explanation of what exactly awareness is or why it might have evolved. It merely postulates that it exists. These two potential views of consciousness have their counterparts in the literature on social perception.

There are presently two main rival views of social perception. The first view, reviewed in the sections above, is that social perception depends on expert systems probably focused on the right TPJ and STS that evolved to compute useful, predictive models of minds. The second view is simulation theory. In simulation theory, social perception is the result of empathy. We understand other people's minds by reference to our own internal experience. The hypothesis about consciousness proposed here was discussed above almost entirely in the framework of the "expert systems" view of social perception. Can the hypothesis be brought into some compatibility with simulation theory? In the present section it is argued that the expert systems view and simulation theory are not mutually exclusive and could at least in principle operate cooperatively to allow for social perception. The hybrid of the two mechanisms is consistent with the present hypothesis about consciousness.

The experimental heart of simulation theory is the phenomenon of mirror neurons. Rizzolatti and colleagues first described mirror neurons in the premotor cortex of macaque monkeys, in a region thought to be involved in the control of the hand for grasping (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Each mirror neuron became active during a particular, complex type of grasp such as a precision grip or a power grip. The neuron also became active when the monkey viewed an experimenter performing the same type of grip. Mirror neurons were therefore both motor and sensory. They responded whether the monkey performed or saw a particular action.

Mirror-like properties were reported in the human cortex in fMRI experiments in which the same area of cortex became active whether the subject performed or saw a particular action (Buccino et al., 2001; Filimon et al., 2007; Iacoboni et al., 2005). The hypothesized mirror-neuron network includes sensory-motor areas of the parietal lobe, anatomically connected regions of the premotor cortex, and perhaps regions of the STS (Rizzolatti & Sinigaglia, 2010).

The hypothesized role of mirror neurons is to aid in understanding the actions of others. In that hypothesis, we understand someone's hand actions by activating our own motor machinery and covertly simulating the actions. The mirror-neuron hypothesis is in some ways an elaboration of Liberman's original hypothesis about speech comprehension (Liberman et al., 1967), in which we categorize speech sounds by using our own motor machinery to covertly mimic the same sounds.

The concept of mirroring can in principle be generalized beyond the perception of other people's hand actions to all social perception. We may understand someone else's joy, complete with nuances and psychological implications, by using our own emotional machinery to simulate that joy. We may understand someone else's intellectual point of view by activating a version of that point of view in our own brains. We may understand other minds in general by simulating them using the same machinery within our own brains. The extent to which mirror neurons directly cause social perception, however, or are a product of more general associative or predictive mechanisms, has been the subject of some discussion (e.g., Heyes, 2010; Kilner et al., 2007).

One difficulty facing simulation theory is that it does not provide any obvious way to distinguish one's own thoughts, intentions, and emotions from someone else's. Both are run on the same hardware. If simulation theory is strictly true, then one should be unable to tell the difference between one's own inner experience and one's perception of someone else's.



A second difficulty facing simulation theory is that it contains some circularity. Before Brain A can mirror the state of Brain B, it needs to know what state to mirror. Brain A needs a mechanism that generates hypotheses about the state of Brain B.

Both of these difficulties with simulation theory disappear in a hybrid system proposed here. The system diagramed in Figure 3 contains expert systems (including TPJ and STS) that contribute to generating models of other minds, and mirror-neuron networks that simulate and thereby refine those models.

As discussed in previous sections, a perceptual model of a mind includes a spatial location assigned to the model. It is true perception in the sense that the perceived properties have a perceived location, to form a perceived object. This spatial embodiment allows one to keep track of whether it is a model of one's own mind or of this or that other person's mind, solving the first difficulty of simulation theory. You know whether the perceived mental states are yours or someone else's in the same way that you know whether a certain color belongs to one object or another—by spatial tag.

A perceptual model of a mind can also supply the necessary information to drive the mirror-neuron simulations. The mirror-neuron system, in this proposal, knows to simulate a reach because the biological motion detectors in the STS have used visual cues to categorize the other person's action as a reach. This process solves the second difficulty of simulation theory. The likely dependence of mirror neurons on an interaction with the STS has been emphasized before (Rizzolatti & Sinigaglia, 2010).

Mirror neurons therefore should not be viewed as rivals to the theory that social perception is emphasized in specialized regions such as the TPJ and STS. Instead, the two mechanisms for social perception could in principle operate in a cooperative fashion. Areas such as TPJ and STS could generate hypotheses about the mind states and intentions of others. Mirror neurons could then use these generated hypotheses to drive simulations. The simulations have the potential to provide a detailed, high quality feedback, resulting in a more elaborate, more accurate model of the other mind. In this proposed scheme (Figure 3), the mirror neuron system is an extended loop adding to and enhancing the machinery that constructs models of minds.

Think how much more complicated, in a recursive, loop-the-loop way, the system becomes when the process of social perception is turned inward. Suppose your self model includes the hypothesis that you are happy right now. To enhance that hypothesis, to enrich the details through simulation, the machinery that constructs your self model contacts and uses your emotion-generating machinery. In that case, if you weren't actually happy, the mirroring process should make you become so as a side product. If you were already happy, perhaps you become more so. Your self model, and the self that is being modeled, become intertwined in complicated ways. Perceiving your own mind changes the thing being perceived, a phenomenon long known to psychologists (Bandura, 2001; Beauregard, 2007).

## Consciousness as information

Figure 4 diagrams a traditional way to conceptualize the problem of consciousness. When information is processed in the brain in some specific but as yet undetermined way (Box 1), a subjective experience emerges (Box 2).

For example, imagine that you are looking at a green apple. Your visual system computes information about the absorption spectrum of the apple. The presence of this wavelength information in the brain can be measured directly by inserting electrodes into visual areas such as cortical area V4. As a result of that information, for unknown reasons, you have a

conscious experience of greenness. We could say that two items are relevant to the discussion: the computation that the apple is green (Box 1 in Figure 4), and the “experienteness” of the green (Box 2)

In a religious or spiritual view, the “experienteness,” the consciousness itself, is a non-physical substance, something like plasma. It is ectoplasm. It is spirit. In new age thinking, it is energy or life force. In traditional Chinese medicine, it is Chi. On introspection people describe it as a feeling, a sense, an experience, a vividness, a private awareness hovering inside the body. In the view of Descartes (1641), it is *res cogitans* or “mental substance.” In the view of the eighteenth century physician Mesmer and the many practitioners who subscribed to his ideas for more than a century, it is a special force of nature called animal magnetism (Alvarado, 2009). In the view of Kant (1781), it is fundamentally not understandable. In the view of Searle (2007), it is like the liquidity of water; it is a state of a thing, not a physical thing. According to many scientists, whatever it is, and however it is caused, it can only happen to information that has entered into a complex, bound unit (Baars, 1983; Crick & Koch, 1990; Engel & Singer, 2001; Tononi, 2008).

Arrow A in Figure 4 represents the process by which the brain generates consciousness. Arrow A is the central mystery to which scientists of consciousness have addressed themselves, with no definite answer or common agreement. It is exceedingly difficult to figure out how a physical machine could produce a non-physical feeling. Our inability to conceive of a route from physical process to mental experience is the reason for the persistent tradition of pessimism in the study of consciousness. When Descartes (1641) claims that *res extensa* (physical substance) can never be used to construct *res cogitans* (mental substance), when Kant (1781) indicates that consciousness can never be understood by reason, when Creutzfeldt (1987) argues that science cannot give an explanation to consciousness, and when Chalmers (1995) euphemistically calls it the “hard problem,” all of these pessimistic views derive from the sheer human inability to imagine how any Arrow A could possibly get from Box 1 to Box 2.

It is instructive, however, to focus instead on Arrow B, a process that is under-emphasized both scientifically and philosophically. Arrow B represents the process by which consciousness can impact the information processing systems of the brain, allowing people to report on the presence of consciousness. We suggest that much more can be learned about consciousness by considering Arrow B than by considering Arrow A. By asking what, specifically, consciousness can *do* in the world, what it can affect, what it can cause, one gains the leverage of objectivity.

Whatever consciousness is, it can ultimately affect speech, since we can talk about it. Here some clarification is useful. Much of the work on consciousness has focused on the information of which one becomes conscious. You can report that you are conscious of this or that. But you can also report on the consciousness itself. You can state confidently that you have an inner feeling, an essence, an awareness, that is attached to this or that item. The essence itself can be reported. If it can be reported, then one must have decided that it is present.

Consider the process of decision-making. Philosophers and scientists of consciousness are used to asking, “What is consciousness?” A more precisely formulated question, relevant to the data that is given to us, would be, “How is it that, on introspection, we consistently decide that we are conscious?” All studies of consciousness, whether philosophical pondering, casual introspection, or formal experiment, depend on a signal-detection and decision-making paradigm. A person answers the question, “Is awareness of X present

inside me?” Note that the question is not, “Is X present,” or, “Is information about X present,” but rather, “Is awareness present?”

Much has been learned recently about the neuronal basis of decision-making, especially in the case of perceptual decisions about visual motion (for review, see Gold & Shadlen, 2007; Sugrue et al., 2005). The decision about visual motion depends on two processes. First, perceptual machinery in the visual system constructs signals that represent motion in particular directions. Second, those signals are received elsewhere in the brain by decision integrators that determine which motion signal is consistent or strong enough to cross a threshold such that a response can be triggered.

Translating the same framework to a decision about awareness, one is led to the following hypothesis. Answering whether you contain awareness depends on at least two processes. First, neuronal machinery (in the present proposal, social perceptual machinery with an emphasis in the right STS and TPJ) generates explicit neuronal signals that represent awareness. Second, a decision process (perhaps in the MPFC) receives and integrates the signals to assess whether the property of awareness is present at a strength above the decision threshold.

The realization that conscious report depends on decision making provides enormous leverage in understanding what, exactly, awareness is. Everything that is known about decision making in the brain points to awareness as explicitly generated neuronal signals that are received by a decision integrator. A similar point has been made with respect to reportable sensory experience by Romo and colleagues (Hernandez et al., 2010; de Lafuente & Romo, 2005).

A crucial property of decision-making is that, not only is the decision itself a manipulation of data, but the decision machine depends on data as input. It does not take any other input. Feeding in some *res cogitans* does not work on this machine. Neither will Chi. You cannot feed it ectoplasm. You cannot feed it an intangible, ineffable, indescribable thing. You cannot feed it an emergent property that transcends information. You can only feed it information. In introspecting, in asking yourself whether you have conscious experience, and in making the decision that yes you have it, what you are deciding on, what you are assessing, the actual stuff your decision engine is collecting and weighing, is information. The experientialness itself, the consciousness, the subjective feeling, the *essence* of awareness that you decide you have, must be information that is generated somewhere in the brain, and that is transmitted to a decision integrator. By considering Arrow B and working backward, we arrive at a conclusion: Consciousness must be information, because only information can be grist for a decision and we can decide that we have consciousness.

One might pose a counter-argument. Suppose consciousness is an emergent property of the brain that is not itself information; but it can affect the processing in the brain, and therefore information about it can be neuronally encoded. We have cognitive access to that information, and therefore ultimately the ability to report on the presence of consciousness. What is wrong with this account? In the theory presented here, nothing is wrong with this account. Indeed it exactly matches the present proposal except in its labeling. In the present theory, the brain does contain an emergent property that is not itself an informational representation. That property is attention. The brain also does contain an informational representation of that property. We can cognitively access that informational representation, thereby allowing us to decide that we have it and report that we have it. However, the item that we introspectively decide that we have and report that we have is not attention; strictly speaking it is the informational representation of attention. In circumstances when the informational representation differs from the thing it represents, we necessarily report the

properties of the representation, because it is that to which we have access. The mysterious and semi-magical properties that we report, that we ascribe to consciousness, must be attributed to the informational representation.

As an analogy, consider an ordinary book that contains, printed in it, the words, “This is not a book. It is really a fire-breathing dragon.” Some readers might announce the presence of a metaphysical mystery. “It seems to be made of simple material in its outer manifestation,” they might say, “but in its inner manifestation, it is something else. How can it be a book and a dragon at the same time? How can we resolve the dualistic mystery between *res libris* and *res dragonis*?” The resolution is simple. It *is* a book. The book contains information on itself. That information is inaccurate.

In the same way, humans have strong intuitions about consciousness, awareness, soul, mind. We look introspectively, that is to say we consult the information present in the brain that describes its own nature, and we arrive at an intuitive understanding in which awareness is a non-physical thing, something like energy or plasma, that has a location vaguely inside the body but no clear physical substance, that feels like something, that is vivid in some cases and less vivid in other cases, that is a private essence, that is experience, that seizes on information. What we are doing, when introspecting in this way, is reading the information in the book. That information is not literally, physically accurate. In the present hypothesis it is a useful informational representation, a depiction of the real physical process of attention. Attention is something the brain does, not something it knows. It is procedural and not declarative. Awareness, in the present hypothesis, is declarative. It is an informational representation that depicts, usefully if not entirely accurately, the process of attention.

We end with one final analogy. Recall Magritte’s famous painting of a pipe with the words scrawled beneath it, “Ceci n’est pas une pipe.” This is not a pipe. It is a representation of a pipe, an existentially deep realization. A distinction exists between a representation and the thing being represented. In the present proposal, consciousness is a representation of attention. But the representation has taken on a life of its own. As we examine it cognitively, abstract properties from it, and verbalize those properties, we find that the representation depicts magic, spirit, soul, inner essence. Since we cannot imagine how those things can be produced physically, we are left with a conundrum. How does the brain produce consciousness? Yet the brain can easily construct information, and information can depict anything, even things that are physically impossible.

## Acknowledgments

Our thanks to the many colleagues who took the time to read drafts and provide invaluable feedback.

## References

- Aflalo TN, Graziano MSA. The organization of the macaque extrastriate visual cortex re-examined using the principle of spatial continuity of function. *Journal of Neurophysiology*. 2011; 105:305–320. [PubMed: 21068269]
- Alvarado CS. Late 19th- and early 20th-century discussions of animal magnetism. *International Journal of Clinical and Experimental Hypnosis*. 2009; 57:366–381. [PubMed: 20182996]
- Apperly IA, Samson D, Chiavarino C, Humphreys GW. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J Cogn Neurosci*. 2004; 16:1773–1784. [PubMed: 15701227]
- Astafiev SV, Shulman GL, Corbetta M. Visuospatial reorienting signals in the human temporo-parietal junction are independent of response selection. *Eur J Neurosci*. 2006; 23:591–596. [PubMed: 16420468]

- Baars, BJ. Conscious contents provide the nervous system with coherent, global information. In: Davidson, RJ.; Schwartz, GE.; Shapiro, D., editors. *Consciousness and Self Regulation*. Vol. 41. NY: Plenum Press; 1983.
- Bandura A. Social cognitive theory: an agentic perspective. *Annu Rev Psychol*. 2001; 52:1–26. [PubMed: 11148297]
- Barracough NE, Xiao D, Oram MW, Perrett DI. The sensitivity of primate STS neurons to walking sequences and to the degree of articulation in static images. *Prog Brain Res*. 2006; 154:135–148. [PubMed: 17010707]
- Baumeister RF, Masicampo EJ. Conscious Thought Is for Facilitating Social and Cultural Interactions: How Mental Simulations Serve the Animal-Culture Interface. *Psychological Review*. 2010; 117:945–971. [PubMed: 20658859]
- Beauregard M. Mind does really matter: evidence from neuroimaging studies of emotional self-regulation, psychotherapy, and placebo effect. *Prog Neurobiol*. 2007; 81:218–236. [PubMed: 17349730]
- Beck DM, Kastner S. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res*. 2009; 49:1154–1165. [PubMed: 18694779]
- Bird CM, Castelli F, Malik O, Frith U, Husain M. The impact of extensive medial frontal lobe damage on ‘Theory of Mind’ and cognition. *Brain*. 2004; 127:914–928. [PubMed: 14998913]
- Birmingham E, Kingstone A. Human social attention: A new look at past, present, and future investigations. *Annals of the New York Academy of Sciences*. 2009; 1156:118–140. [PubMed: 19338506]
- Blakemore SJ, Boyer P, Pachot-Clouard M, Meltzoff A, Segebarth C, Decety J. The detection of contingency and animacy from simple animations in the human brain. *Cerebral Cortex*. 2003; 13:837–844. [PubMed: 12853370]
- Blanke O, Mohr C, Michel CM, Pascual-Leone A, Brugger P, Seeck M, Landis T, Thut G. Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J Neurosci*. 2005; 25:550–557. [PubMed: 15659590]
- Blanke O, Ortigue S, Landis T, Seeck M. Stimulating illusory own-body perceptions. *Nature*. 2002; 419:269–270. [PubMed: 12239558]
- Block N. How can we find the neural correlates of consciousness? *Trends in Neurosciences*. 1996; 19:456–459. [PubMed: 8931266]
- Brain WR. A Form of Visual Disorientation Resulting from Lesions of the Right Cerebral Hemisphere. *Proc R Soc Med*. 1941; 34:771–776. [PubMed: 19992404]
- Bruce C, Desimone R, Gross CG. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol*. 1981; 46:369–384. [PubMed: 6267219]
- Brunet E, Sarfati Y, Hardy-Baylé MC, Decety J. A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage*. 2000; 11:157–166. [PubMed: 10679187]
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci*. 2001; 13:400–404. [PubMed: 11168545]
- Calder AJ, Lawrence AD, Keane J, Scott SK, Owen AM, Christoffels I, Young AW. Reading the mind from eye gaze. *Neuropsychologia*. 2002; 40:1129–1138. [PubMed: 11931917]
- Carruthers P. How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*. 2009; 32:121–182. [PubMed: 19386144]
- Chalmers D. Facing up to the problem of consciousness. *J Consciousness Studies*. 1995; 2:200–219.
- Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, Walter H. The intentional network: how the brain reads varieties of intentions. *Neuropsychologia*. 2007; 45:3105–3113. [PubMed: 17669444]
- Coover JE. The feeling of being stared at. *Am. J. Psych*. 1913; 24:570–575.
- Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci*. 2000; 3:292–297. [PubMed: 10700263]

- Cottrell JE, Winer GA. Development in the understanding of perception: The decline of extromission perception beliefs. *Dev Psych*. 1994; 30:218–228.
- Creutzfeldt, O. Inevitable deadlocks of the brain-mind discussion. In: Gulayas, B., editor. *The Brain Mind Problem: Philosophical and Neurophysiological Approaches*. Leuven University Press; 1987. p. 3-27.
- Crick F, Koch C. Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences*. 1990; 2:263–275.
- Critchley, M. *The Parietal Lobes*. London: Hafner Press; 1953.
- Damasio AR. Synchronous activation in multiple cortical regions: a mechanism for recall. *Seminars in neuroscience*. 1990; 2:287–296.
- Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci*. 2006; 10:204–211. [PubMed: 16603406]
- de Lafuente V, Romo R. Neuronal correlates of subjective sensory experience. *Nature neuroscience*. 2005; 8:1698–1703.
- Descartes, R. *Meditations on First Philosophy*, in *The Philosophical Writings of René Descartes*. Cottingham, J.; Stoothoff, R.; Murdoch, D., editors. Vol. vol. 2. Cambridge: Cambridge University Press; 1641. p. 1-62.1984
- Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci*. 1984; 4:2051–2062. [PubMed: 6470767]
- Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci*. 1995; 18:193–222. [PubMed: 7605061]
- di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. *Exp Brain Res*. 1992; 1992; 91:176–180. [PubMed: 1301372]
- Ehrsson HH. The experimental induction of out-of-body experiences. *Science*. 2007; 317:1048. [PubMed: 17717177]
- Engel AK, König P, Gray CM, Singer W. Stimulus-Dependent Neuronal Oscillations in Cat Visual Cortex: Inter-Columnar Interaction as Determined by Cross-Correlation Analysis. *Eur J Neurosci*. 1990; 2:588–606. [PubMed: 12106294]
- Engel AK, Singer W. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci*. 2001; 5:16–25. [PubMed: 11164732]
- Ferber S, Karnath HO. How to assess spatial neglect-line bisection or cancellation tasks. *J Clin Exp Neuropsychol*. 2001; 23:599–607. [PubMed: 11778637]
- Filimon F, Nelson JD, Hagler DJ, Sereno MI. Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage*. 2007; 37:1315–1328. [PubMed: 17689268]
- Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition*. 1995; 57:109–128. [PubMed: 8556839]
- Friesen CK, Kingstone A. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychon. Bull Rev*. 1998; 5:490–495.
- Frischen A, Bayliss AP, Tipper SP. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull*. 2007; 133:694–724.
- Frith C. Consciousness is for other people. *Behavioral and Brain sciences*. 1995; 18:682–683.
- Frith C. Attention to action and awareness of other minds. *Consciousness and Cognition*. 2002; 11:481–487. [PubMed: 12470618]
- Frith U, Frith CD. Development and neurophysiology of mentalizing. *Philos Trans R Soc Lond B Biol Sci*. 2003; 358:459–473. [PubMed: 12689373]
- Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD. Reading the mind in cartoons and stories: an fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia*. 2000; 38:11–21. [PubMed: 10617288]
- Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. *Brain*. 1996; 119:593–609. [PubMed: 8800951]
- Gazzaniga MS. *The Bisected Brain*. Appleton Century Crofts. 1970

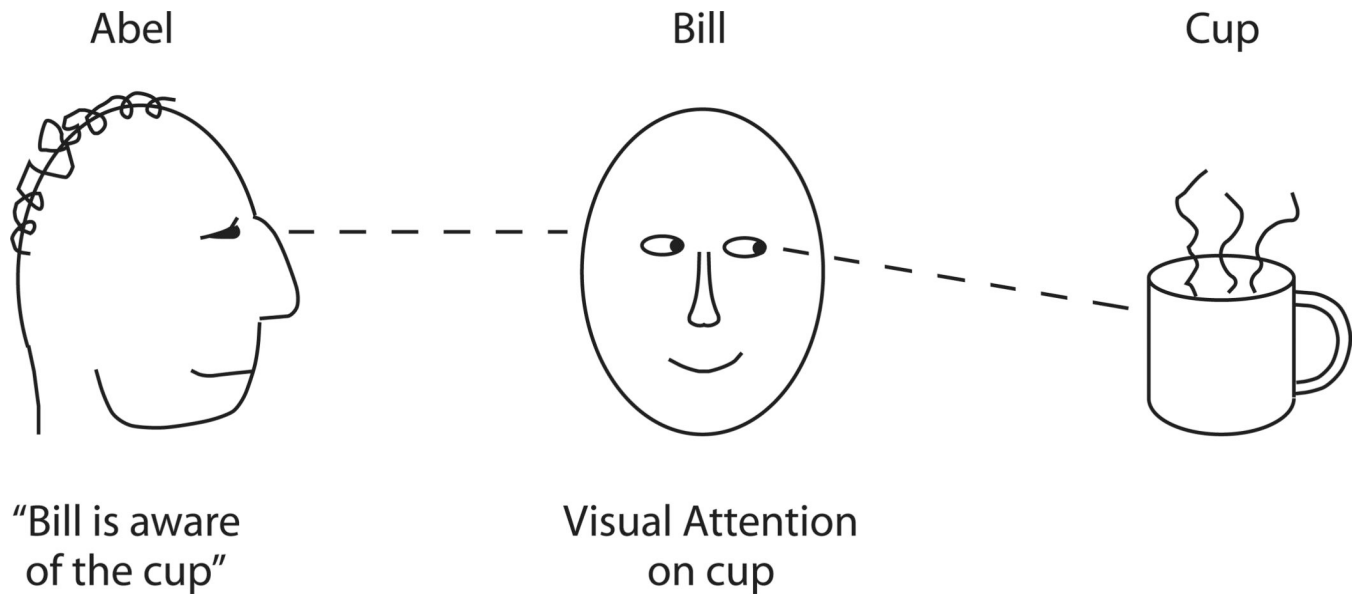
- Goel V, Grafman J, Sadato N, Hallett M. Modeling other minds. *Neuroreport*. 1995; 6:1741–1746. [PubMed: 8541472]
- Gold JI, Shadlen MN. The neural basis of decision making. *Annu Rev Neurosci*. 2007; 30:535–574. [PubMed: 17600525]
- Graziano, MSA. *God, Soul, Mind, Brain: A Neuroscientist's Reflections on the Spirit World*. Teaticket MA: Leapfrog Press; 2010.
- Graziano MS, Aflalo TN. Rethinking cortical organization: moving away from discrete areas arranged in hierarchies. *Neuroscientist*. 2007; 13:138–147. [PubMed: 17404374]
- Graziano, MSA.; Botvinick, MM. How the brain represents the body: insights from neurophysiology and psychology. In: Prinz, W.; Hommel, B., editors. *Common Mechanisms in Perception and Action: Attention and Performance XIX*. Oxford England: Oxford University Press; 2002. p. 136-157.
- Gross CG. The fire that comes from the eye. *The Neuroscientist*. 1999; 5:58–64.
- Gross CG, Bender DB, Rocha-Miranda CE. Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*. 1969; 166:1303–1306. [PubMed: 4982685]
- Gross CG, Graziano MSA. Multiple representations of space in the brain. *The Neuroscientist*. 1995; 1:43–50.
- Grossberg S. The link between brain learning, attention, and consciousness. *Consciousness and Cognition*. 1999; 8:1–44. [PubMed: 10072692]
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. Brain areas involved in perception of biological motion. *J Cogn Neurosci*. 2000; 12:711–720. [PubMed: 11054914]
- Halligan PW, Fink GR, Marshall JC, Vallar G. Spatial cognition: evidence from visual neglect. *Trends Cogn Sci*. 2003; 7:125–133. [PubMed: 12639694]
- Halligan PW, Marshall JC. Left visuo-spatial neglect: a meaningless entity? *Cortex*. 1992; 28:525–535. [PubMed: 1478083]
- Heilman KM, Valenstein E. Frontal lobe neglect in man. *Neurology*. 1972a; 22:660–664. [PubMed: 4673341]
- Heilman KM, Valenstein E. Mechanism underlying hemispatial neglect. *Ann Neurol*. 1972b; 5:166–170. [PubMed: 426480]
- Hernandez A, Nacher V, Luna R, Zainos A, Lemus L, Alvarez M, Vazquez Y, Camarillo L, Romo R. Decoding a Perceptual Decision Process across Cortex. *Neuron*. 2010; 66:300–314. [PubMed: 20435005]
- Heyes C. Where do mirror neurons come from? *Neurosci Biobehav Rev*. 2010; 34:575–583. [PubMed: 19914284]
- Hoffman EA, Haxby JV. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci*. 2000; 3:80–84. [PubMed: 10607399]
- Humphrey, N. *Consciousness Regained: Chapters in the Development of Mind*. Oxford: Oxford University Press; 1983.
- Iacoboni M, Molnar-Szakacs I, Gallese V, Buccino G, Mazziotta JC, Rizzolatti G. Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol*. 2005; 3:79.
- Jiang Y, Costello P, Fang F, Huang M, He S. A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proc Natl Acad Sci U S A*. 2006; 103:17048–17052. [PubMed: 17075055]
- Jellema T, Perrett DI. Cells in monkey STS responsive to articulated body motions and consequent static posture: a case of implied motion? *Neuropsychologia*. 2003; 41:1728–1737. [PubMed: 14527537]
- Jellema T, Perrett DI. Neural representations of perceived bodily actions using a categorical frame of reference *Neuropsychologia*. 2006; 44:1535–1546.
- Kant, I. *Critik der reinen Vernunft*. English translation by F. M. Muller. New York: Anchor Books; 1781. 1966
- Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 1997; 17:4302–4311. [PubMed: 9151747]

- Karnath HO, Ferber S, Himmelbach M. Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*. 2001; 411:950–953. [PubMed: 11418859]
- Kentridge RW, Heywood CA, Weiskrantz L. Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia*. 2004; 42:831–835. [PubMed: 15037061]
- Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. *Cogn Process*. 2007; 8:159–566. [PubMed: 17429704]
- Kinsbourne M. A model for the mechanism of unilateral neglect of space. *Trans Am Neurol Assoc*. 1970; 95:143–146. [PubMed: 5514359]
- Koch C, Tsuchiya N. Attention and consciousness: two distinct brain processes. *Trends Cogn Sci*. 2007; 11:16–22. [PubMed: 17129748]
- Lamme VA. Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Netw*. 2004; 17:861–872. [PubMed: 15288903]
- Lamme VA. Towards a true neural stance on consciousness. *Trends Cogn Sci*. 2006; 10:494–501. [PubMed: 16997611]
- Lenggenhager B, Tadi T, Metzinger T, Blanke O. Video ergo sum: manipulating bodily self-consciousness. *Science*. 2007; 317:1096–1099. [PubMed: 17717189]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev*. 1967; 74:431–461. [PubMed: 4170865]
- Linias, R.; Ribary, U. Perception as an oneiric-like state modulated by the senses. In: Koch, C.; Davis, J., editors. *Large Scale Neural Theories of the Brain*. Cambridge USA: MIT Press; 1994. p. 111-124.
- Meister IG, Wienemann M, Buelte D, Grünewald C, Sparing R, Dambeck N, Boroojerdi B. Hemiextinction induced by transcranial magnetic stimulation over the right temporo-parietal junction. *Neuroscience*. 2006; 142:119–123. [PubMed: 16876326]
- Mesulam MM. A cortical network for directed attention and unilateral neglect. *Ann Neurol*. 1981; 10:309–325. [PubMed: 7032417]
- Mesulam MM. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos Trans R Soc Lond B Biol Sci*. 1999; 354:1325–1346. [PubMed: 10466154]
- Mitchell LP. Activity in the right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*. 2008; 18:262–271. [PubMed: 17551089]
- Mort DJ, Malhotra P, Mannan SK, Rorden C, Pambakian A, Kennard C, Husain M. The anatomy of visual neglect. *Brain*. 2003; 126:1986–1997. [PubMed: 12821519]
- Naccache L, Blandin E, Dehaene S. Unconscious masked priming depends on temporal attention. *Psychol Sci*. 2002; 13:416–424. [PubMed: 12219807]
- Newman J, Baars BJ. A neural attentional model for access to consciousness: a global workspace perspective. *Concepts in Neuroscience*. 1993; 4:255–290.
- Nisbett RE, Wilson TD. Telling More Than We Can Know - Verbal Reports on Mental Processes. *Psychological Review*. 1977; 84:231–259.
- Nummenmaa L, Calder AJ. Neural mechanisms of social attention. *Trends in Cognitive Sciences*. 2008; 13:135–143. [PubMed: 19223221]
- Ochsner KN, Knierim K, Ludlow DH, Hanelin J, Ramchandran T, Glover G, Mackey SC. Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J Cogn Neurosci*. 2004; 16:1746–1772. [PubMed: 15701226]
- Pasquali A, Timmermans B, Cleeremans A. Know thyself: Metacognitive networks and measures of consciousness. *Cognition*. 2010; 117:182–190. [PubMed: 20825936]
- Passingham RE, Bengtsson SL, Lau HC. Medial frontal cortex: from self-generated action to reflection on one's own performance. *Trends Cogn Sci*. 2010; 14:16–21. [PubMed: 19969501]
- Pelphrey KA, Morris JP, McCarthy G. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci*. 2004; 16:1706–1716. [PubMed: 15701223]



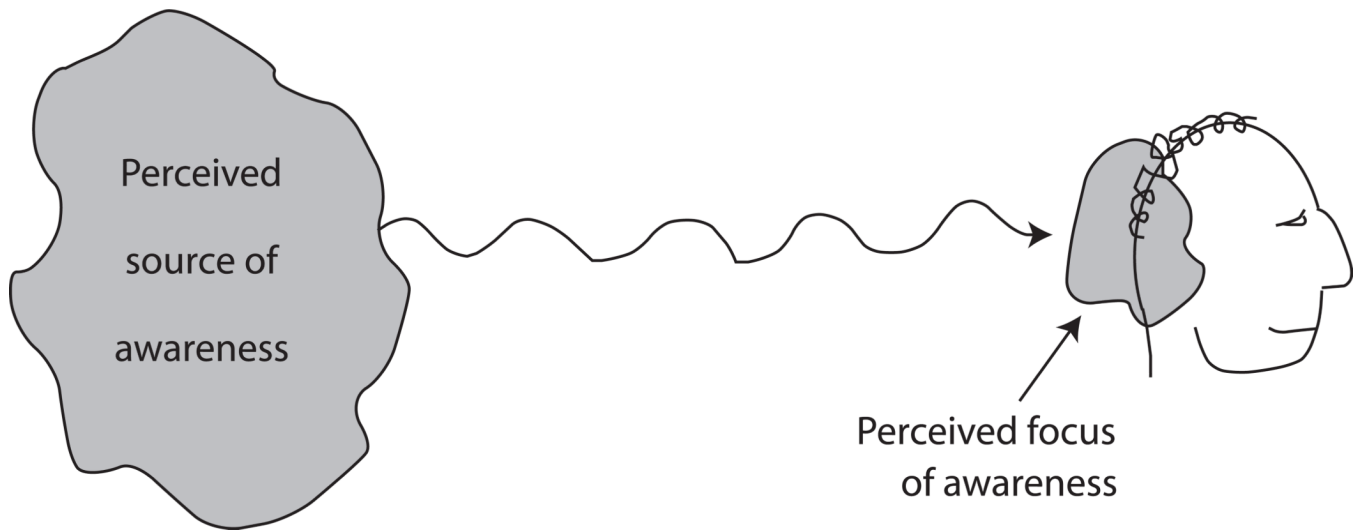
- Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G. Functional anatomy of biological motion perception in posterior temporal cortex: an FMRI study of eye, mouth and hand movements. *Cereb Cortex*. 2005; 15:1866–1876. [PubMed: 15746001]
- Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA. Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc Lond B Biol Sci*. 1985; 223:293–317. [PubMed: 2858100]
- Premack D, Woodruff G. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*. 1978; 1:515–526.
- Ptak R, Schnider A. The dorsal attention network mediates orienting toward behaviorally relevant stimuli in spatial neglect. *J Neurosci*. 2010; 30:12557–12565. [PubMed: 20861361]
- Puce A, Allison T, Bentin S, Gore JC, McCarthy G. Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci*. 1998; 18:2188–2199. [PubMed: 9482803]
- Rizzolatti G, Fadiga L, Gallese V, Fogassi L. Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res*. 1996; 3:131–141. [PubMed: 8713554]
- Rizzolatti G, Sinigaglia C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci*. 2010; 11:264–274. [PubMed: 20216547]
- Rorden C, Fruhmann Berger M, Karnath HO. Disturbed line bisection is associated with posterior brain lesions. *Brain Res*. 2006; 1080:17–25. [PubMed: 16519881]
- Rosenthal DM. Consciousness, content, and metacognitive judgments. *Consciousness and Cognition*. 2000; 9:203–214. [PubMed: 10924239]
- Samson D, Apperly IA, Braithwaite JJ, Andrews BJ, BodleyScott SE. Seeing it their way: evidence for rapid and involuntary computation of what other people see. *J Exp Psychol Hum Percept Perform*. 2010; 36:1255–1266. [PubMed: 20731512]
- Samson D, Apperly IA, Chiavarino C, Humphreys GW. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci*. 2004; 7:499–500. [PubMed: 15077111]
- Saxe R, Kanwisher N. People thinking about thinking people: fMRI investigations of theory of mind. *NeuroImage*. 2003; 19:1835–1842. [PubMed: 12948738]
- Saxes R, Moran JM, Scholz J, Gabrieli J. Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Soc Cogn Affect Neurosci*. 2006; 1:229–234. [PubMed: 18985110]
- Saxe R, Wexler A. Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*. 2005; 43:1391–1399. [PubMed: 15936784]
- Scholz J, Triantafyllou C, Whitfield-Gabrieli S, Brown EN, Saxe R. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One*. e4869; 4
- Searle JR. Dualism revisited. *J Physiol Paris*. 2007; 101:169–178. [PubMed: 18276124]
- Shulman GL, Pope DL, Astafiev SV, McAvoy MP, Snyder AZ, Corbetta M. Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *J Neurosci*. 2010; 30:3640–3651. [PubMed: 20219998]
- Sugrue LP, Corrado GS, Newsome WT. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat Rev Neurosci*. 2005; 6:363–375. [PubMed: 15832198]
- Szczepanski SM, Konen CS, Kastner S. Mechanisms of spatial attention control in frontal and parietal cortex. *J Neurosci*. 2010; 30:148–160. [PubMed: 20053897]
- Titchner EB. The feeling of being stared at. *Science*. 1898; 8:895–897.
- Thompson JC, Hardee JE, Panayiotou A, Crewther D, Puce A. Common and distinct brain activation to viewing dynamic sequences of face and hand movements. *Neuroimage*. 2007; 37:966–973. [PubMed: 17616403]
- Tononi G. Consciousness as integrated information: a provisional manifesto. *Biological Bulletin*. 2008; 215:216–242. [PubMed: 19098144]
- Tononi G, Edelman GM. Consciousness and Complexity. *Science*. 1998; 282:1846–1851. [PubMed: 9836628]
- Treisman A. Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*. 1988; 40A:201–237. [PubMed: 3406448]

- Treisman A, Gelade G. A feature-integration theory of attention. *Cognitive Psychology*. 1980; 12:97–136. [PubMed: 7351125]
- Vaina LM, Solomon J, Chowdhury S, Sinha P, Belliveau JW. Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci USA*. 2001; 98:11656–11661. [PubMed: 11553776]
- Vallar G. Extrapersonal visual unilateral spatial neglect and its neuroanatomy. *Neuroimage*. 2001; 14:S52–S58. [PubMed: 11373133]
- Vallar G, Perani D. The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia*. 1986; 24:609–622. [PubMed: 3785649]
- Vogeley K, Bussfeld P, Newen A, Herrmann S, Happé F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K. Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage*. 2001; 14:170–181. [PubMed: 11525326]
- Vogeley K, May M, Ritzl A, Falkai P, Zilles K, Fink GR. Neural correlates of first-person perspective as one constituent of human self-consciousness. *J Cogn Neurosci*. 2004; 16:817–827. [PubMed: 15200709]
- Von der Malsburg, C. The coherence definition of consciousness. In: Ito, M.; Miyashita, Y.; Rolls, E., editors. *Cognition, Computation, and Consciousness*. Oxford: Oxford University Press; 1997. p. 193-204.
- Weed E, McGregor W, Feldbaek Nielsen J, Roepstorff A, Frith U. Theory of Mind in adults with right hemisphere damage: What's the story? *Brain Lang*. 2010; 113:65–72. [PubMed: 20188408]
- Wicker B, Michel F, Henaff MA, Decety J. Brain regions involved in the perception of gaze: a PET study. *Neuroimage*. 1988; 8:221–227. [PubMed: 9740764]
- Wyk BC, Hudac CM, Carter EJ, Sobel DM, Pelphrey KA. Action understanding in the superior temporal sulcus region. *Psychol Sci*. 2009; 20:771–777. [PubMed: 19422619]
- Zacks JM, Vettel JM, Michelon P. Imagined viewer and object rotations dissociated with event-related fMRI. *J Cogn Neurosci*. 2003; 15:1002–1018. [PubMed: 14614811]

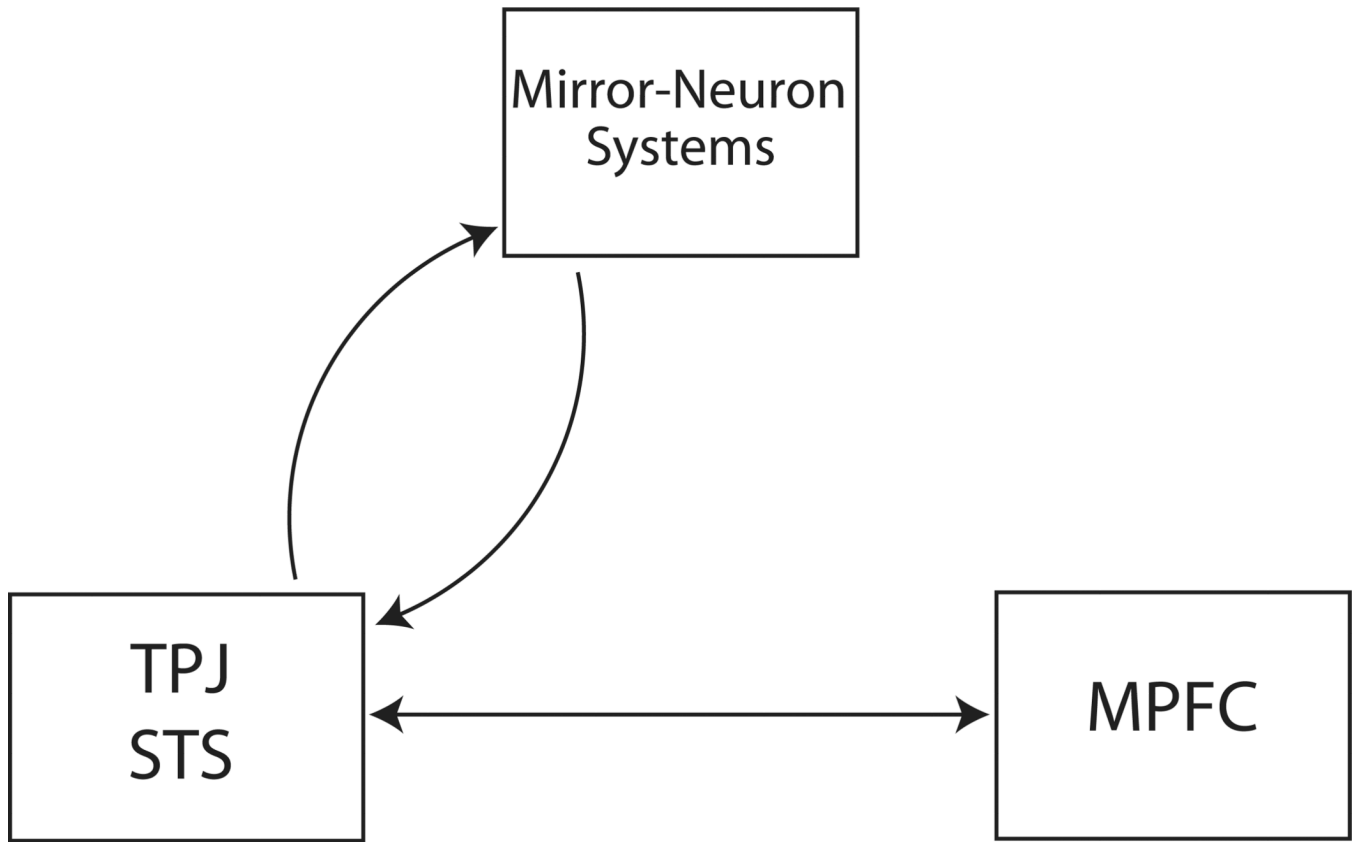


**Figure 1.**

Awareness as a social perceptual model of attention. Bill has his visual attention on the cup. Abel, observing Bill, constructs a model of Bill's mental state using specialized neuronal machinery for social perception. Part of that model is the proposition that Bill is aware of the cup. In this formulation, awareness is a perceptual property that is constructed to represent the attentional state of a brain. We perceive awareness in other people. We can use the same neuronal machinery of social perception to perceive awareness in ourselves.

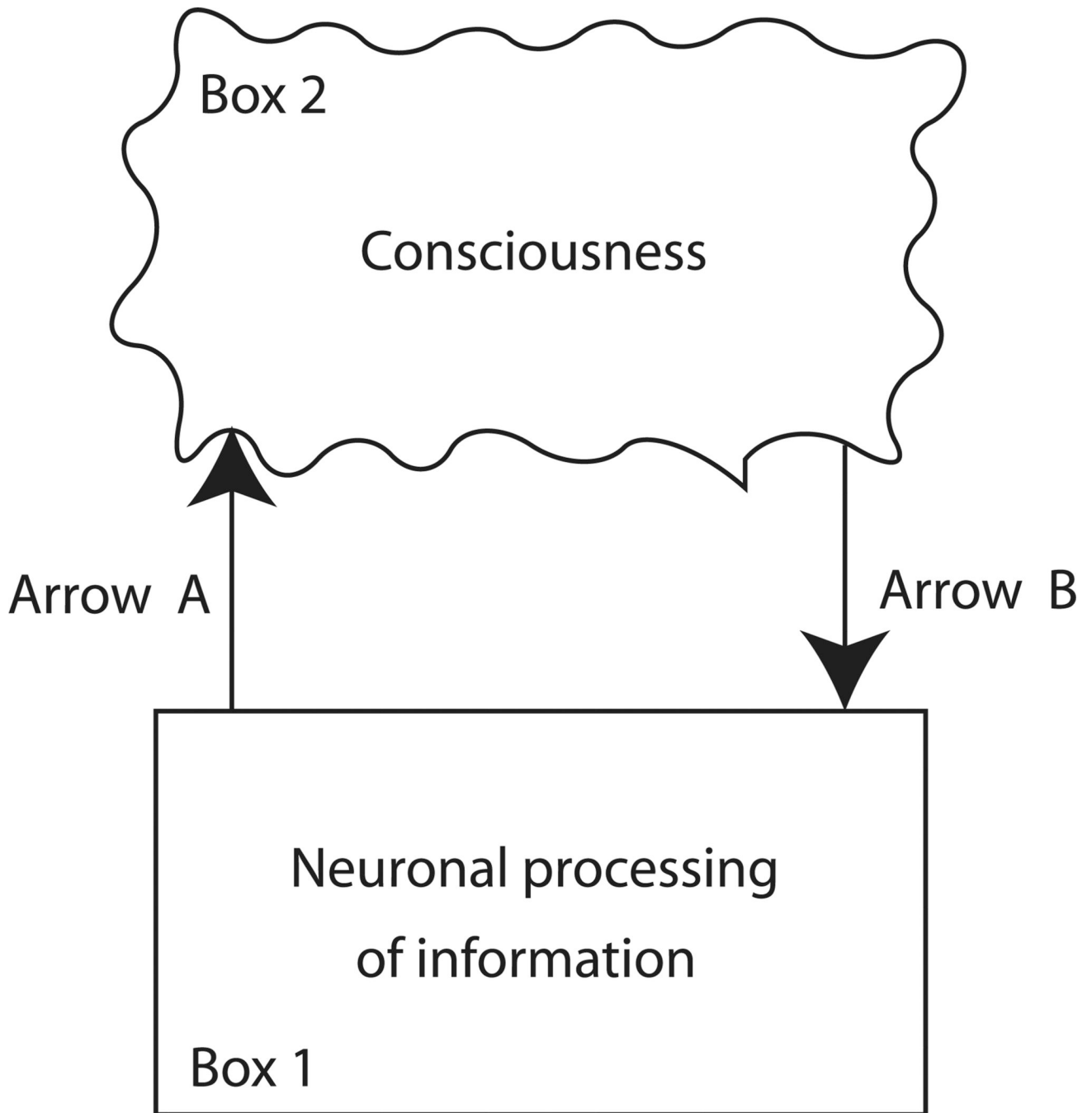


**Figure 2.**  
The perceptual illusion that somebody behind you is staring at you.



**Figure 3.**

A proposed scheme that integrates simulation theory with the theory of dedicated cortical areas for social cognition. The box labeled “TPJ, STS” represents a cluster of cortical areas that contributes to building perceptual models of minds, including a self-model and models of other minds. The box labeled “MPFC” represents a prefrontal area that contributes to decisions in the social domain. The box labeled “Mirror-Neuron Systems” represents brain-wide networks that simulate and thereby refine the models of minds generated in TPJ and STS.



**Figure 4.**

A traditional view in which consciousness emerges from the information processed in the brain. Consideration of Arrow A, how the brain creates consciousness, leads to much controversy and little insight. Consideration of Arrow B, how consciousness affects the brain, leads to the inference that consciousness must be information, because only information can act as grist for decision machinery, and we can decide that we have consciousness.