Speculations on the Evolution of Awareness

Michael S. A. Graziano

Abstract

■ The "attention schema" theory provides one possible account of the biological basis of consciousness, tracing the evolution of awareness through steps from the advent of selective signal enhancement about half a billion years ago to the top–down control of attention, to an internal model of attention (which allows a brain, for the first time, to attribute to itself that it has a mind that is aware of something), to the ability to attribute awareness to other beings, and from there to the human attribution of a rich spirit world surrounding us. Humans have been known to attribute awareness to plants, rocks, rivers, empty space, and the universe as a whole. Deities, ghosts, souls—the spirit world swirling around us is arguably the exuberant attribution of awareness. ■

INTRODUCTION

The topic of consciousness has been approached, at least within psychology and neuroscience, from two broad traditional perspectives. First, the content of consciousness can be studied. For example, perhaps consciousness contains a running narrative that is used to explain one's own behavior (e.g., Libet, Gleason, Wright, & Pearl, 1983; Nisbett & Wilson, 1977; Gazzaniga, 1970). This approach tends to emphasize confabulation and attribution and is closely related to social psychology because people can attribute mental properties such as motivations and intentions to themselves and to others.

Second, one can ask, regardless of the content that is within consciousness, how does the content acquire subjective experience (e.g., Tononi, 2008; Chalmers, 1995; Crick & Koch, 1990)? What is awareness itself? This second approach is more typically applied to sensory awareness, such as the awareness of color. The visual system constantly computes information about the colors of surfaces, but only a small amount of that information enters reportable awareness. What is the distinction between information that does not have subjective experience attached to it and information that does?

Recently, my colleagues and I proposed a theory of awareness that draws on both approaches to consciousness (Graziano, 2013; Graziano & Kastner, 2011). The theory addresses a specific question: How and for what possible adaptive advantage do brains attribute the property of subjective experience to some instances of internally computed information and not other instances? This article briefly outlines the "attention schema" theory.

Princeton University

SELECTIVE SIGNAL ENHANCEMENT AND ATTENTION

The proposed theory begins with something much simpler than awareness: the evolution of selective signal enhancement or the ability of neural nets to boost the most useful signals of the moment at the expense of other signals (see Figure 1).

The hydra has a neural net but apparently no clear selective signal enhancement (Bode, Heimfeld, Koizumi, Littlefield, & Yaross, 1988). Selective signal enhancement therefore presumably evolved after the branch between medusozoa and other animals. This branch point is thought to have occurred roughly 550 million years ago (MYA), possibly longer (Budd, 2008). Selective signal enhancement is present in a great range of other animals, such as the crab (Barlow & Fraioli, 1978). Using competitive mechanisms, neurons in the crab eye enhance information about the borders between dark and light. Those enhanced signals can then have a larger impact on behavior. Selective signal enhancement is also present in the fly visual system (van Swinderen, 2012), bird visual system (Mysore & Knudsen, 2013), and primate brain (Beck & Kastner, 2009). These many branches of the animal kingdom share a common ancestor between about 550 and 500 MYA, around the time of the Cambrian explosion. One can therefore hazard an educated guess: About half a billion years ago, neuronal nets evolved a fundamental new ability that allowed salient signals to win a competition and become enhanced at the expense of other signals.

Selective signal enhancement is bottom–up. But attention, especially as it has been studied in primates, has a top–down component. Desimone and colleagues worked out the story of biased competition in the primate brain (Beck & Kastner, 2009; Desimone & Duncan, 1995). Internally generated directives can have a top–down influence,

Journal of Cognitive Neuroscience 26:6, pp. 1300–1304
doi:10.1162/jocn_a_00623

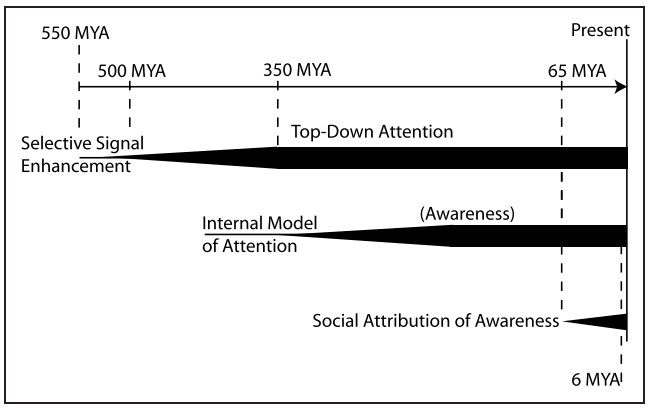

Figure 1. From selective signal enhancement to consciousness. In the present proposal, about half a billion years ago, nervous systems evolved an ability to enhance the most pressing of incoming signals. Gradually, this signal enhancement came under top–down control and became selective attention. To effectively predict and deploy its own attentional focus, the brain may have evolved a constantly updated simulation of attention or attention schema. Instead of attributing a complex neuronal machinery to the self, this schema attributes to the self an experience of X—the property of being aware of something. Just as the brain can direct attention to external signals or to internal signals, this model of attention can attribute to the self an awareness of external events or of internal event. As the model increased in sophistication, it came to be used not only for modeling one's own attention but also for understanding other beings by modeling their possible states of attention. The theory explains why a brain attributes the property of awareness to itself and why we humans are so prone to attribute awareness to the people and objects around us. Timeline: Hydras evolved approximately 550 MYA with no selective signal enhancement. Animals that do show selective signal enhancement diverged from each other between approximately 550 and 500 MYA. Animals such as birds and mammals that show sophisticated top–down control of attention diverged from each other approximately 350 MYA. Primates first appeared approximately 65 MYA. Hominins appeared approximately 6 MYA.

biasing the competition among incoming signals. A sophisticated top–down control of attention has not been studied systematically across phylogenetic clades, but it is known to be highly developed in at least birds and mammals (Mysore & Knudsen, 2013; Beck & Kastner, 2009) that have a common ancestor about 350 MYA. By implication, somewhere in the span from about 550 to 350 MYA, attention as we think of it evolved from a simpler, selective signal enhancement into a sophisticated interplay of bottom–up and top–down mechanisms. Of course it continued to evolve, and most of what we know about it pertains to the primate brain.

AN INTERNAL MODEL OF ATTENTION

One of the principles of control theory is that, to effectively control a complex variable, it is useful for the controller to have an internal model of that variable, including an ability to simulate its dynamics, monitor its state, and predict its state at least a few seconds into the future. To control the movement of the arm, for example, the brain constructs a constantly updated internal model or computational simulation of the arm (Hwang & Shadmehr, 2005; Wolpert, Ghahramani, & Jordan, 1995). Just so, as the brain evolved a top–down control of attention, we hypothesize (Graziano, 2013; Graziano & Kastner, 2011) that it evolved an internal model of attention (see Figure 1).

The brain constructs models of things that are useful to monitor, predict, and control. These models are always simplifications. They are quick-and-dirty because they need to be computed and updated in real time. No model constructed by the brain is ever really accurate. To understand the theory summarized here, it is necessary to understand the enormity of that gap between the model and the thing being modeled. For example, the internal model of the arm is inaccurate in a variety of ways. The actual arm has an internal structure of bones, joints, tendons, fat, blood vessels, and other architecture, none of which is represented in the body schema that is computed in the brain. The body schema is a shell model, a surface model containing a few need-to-know bits of information. It is a physically incoherent model because one cannot physically have the shell of an arm, which moves like an arm, without the structure inside it. Moreover, the model is notoriously easy to fool (Graziano & Botvinick, 2002). Put your hand under a table, and within about half a minute, your sense of arm position begins to drift. Other illusions such as the tendon vibration illusion (Lackner, 1988) and the rubber hand illusion (Botvinick & Cohen, 1998) can easily introduce a discrepancy between the actual state of the arm and the internal model of the arm.

Just so, we proposed that the brain constructs an internal model or simulation of attention, and like all models constructed by the brain, this one is a limited, partial, distorted, and schematic model. What might a model of attention look like? We termed it the *attention schema*. We argue that in humans it has evolved into an idiosyncratic set of properties. When you pay attention to item X, the attention schema models that state. But the model does not depict the details of neurons and competitive interactions. The brain does not know those mechanistic details about itself. Instead the model attributes to yourself the property: I experience X. I have a mind that is occupied by X. I am subjectively aware of X. It is, if you will, a shell model or surface model of attention, a quick and dirty model of the basics without the details of mechanism. The brain attributes to itself an ethereal state of experience, of a mind that can experience, of subjectivity, because that is a good enough cartoon sketch of a brain focusing its attention. With the evolution of this attention schema, brains have an ability to attribute to themselves not only "this object is green" or "I am a living being," but also, "I have a subjective experience of those items."

It has been noted by many researchers that awareness and attention have a complex relationship (e.g., Koch & Tsuchiya, 2007; Lamme, 2004; O'Regan & Noë,

2001; Posner, 1994). Two aspects of that relationship are summarized here.

First, attention is something the brain does. It is a data-handling method in which selected signals are enhanced at the expense of other signals. But awareness is something the brain knows. The brain can decide that it has awareness and can potentially report it. And that is precisely the relationship suggested here: Awareness is a schematic, informational model of something, and attention is the thing being modeled.

A second key aspect of the relationship between attention and awareness is that, although they often covary, they are not the same and can be dissociated (e.g., Tallon-Baudry, 2011; Koch & Tsuchiya, 2007; Kentridge, Heywood, & Weiskrantz, 2004; Lamme, 2004; Naccache, Blandin, & Dehaene, 2002). In particular, it is possible to attend to a visual stimulus, in the sense of focusing processing resources on it, while having no reportable awareness of the stimulus. This dissociation is particularly easy to produce for visual stimuli that are weak or masked and that therefore fall below detection threshold. This dissociation has led to a flurry of speculation about whether attention and awareness have any relationship to each other at all or are entirely independent processes. Yet for a stimulus above detection threshold in a normal brain, if you are attending to it, you are typically also aware of it. Although the two can be dissociated, they typically covary. We suggested (Graziano, 2013; Graziano & Kastner, 2011) that awareness and attention have a specific type of relationship to each other, the relationship between a model and the item being modeled. The model is schematic, rough, sometimes inaccurate, and therefore is sometimes dissociable from the item being modeled.

LARGE-SCALE INTEGRATION OF INFORMATION

Suppose that you are attending to an apple in front of you. How might a brain model this state of attention? As illustrated schematically in Figure 2, the apple is represented partly by the encoded sensory information: color, shape, texture, and also semantic knowledge. In addition, the brain constructs information about yourself: your physical location and body, your mood, your intentions, your self-knowledge. Finally, in the present proposal, a third chunk of information is necessary, a model of what it means for a brain to focus its attention on something, the attention schema. Cognitive machinery, accessing this vast set of interlinked information, can report not only, "There is an I and there is an apple," but, "There is an I who has an awareness of the apple."

It has been suggested that consciousness involves a global workspace of information (Newman & Baars, 1993; Baars, 1983), or an integration of information (Tononi, 2008), or that it depends on neural oscillations or thalamocortical loops that may form the mechanism for binding information across brain areas (Schiff, 2008; Engel & Singer, 2001; Crick & Koch, 1990). These proposals are consistent with the present theory, which depends on a large-scale integration of disparate pieces of information. But the current theory also suggests that the core evolutionary development regarding awareness, the essential chunk of information at the heart of the global workspace, is the internal model of attention, a model that depicts what it means for an agent to be aware of something. Without that model, the brain might still contain a set of integrated information about the world outside and inside but would lack any basis to conclude internally or report externally that an experience or an awareness is related to that content.

SOCIAL USES OF AWARENESS

We suggested that the internal model of attention may be used not only to model one's own internal state of attention but also to monitor and predict the attentional state of others (Graziano, 2013; Graziano & Kastner, 2011).

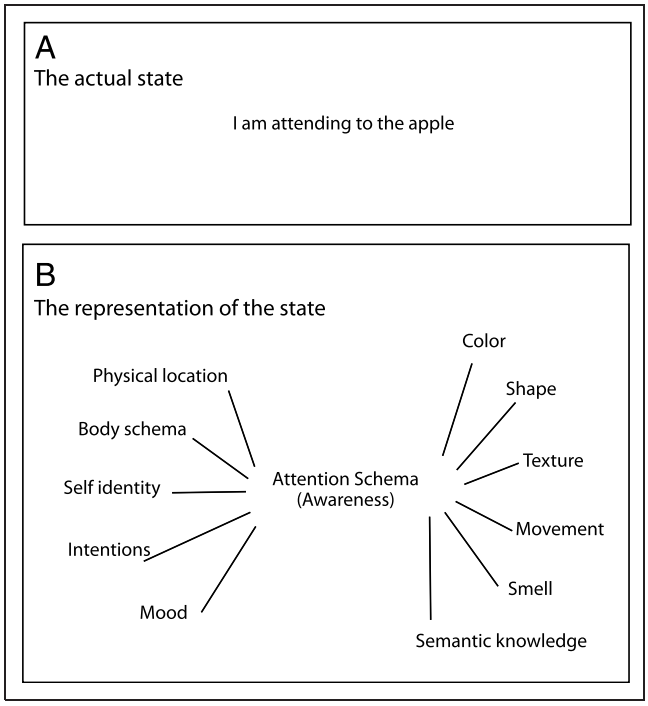

Figure 2. Representing attention. Suppose you are attending to an apple (A). Each part of that condition can be represented by means of information encoded and linked in the brain (B). The apple is encoded by means of many chunks of information encompassing sensory and semantic properties. The self is represented by chunks of information about one's own body, personhood, goals, and emotions. According to the present theory, the attentive relationship between you and the apple is represented by the attention schema, which uses the construct of awareness as a model of attention. These many pieces, bound together, provide a larger informational model or representation. Cognitive machinery can access that representation and summarize its contents. Thus, on introspection, you conclude that there is a self who is aware of the apple. Without the attention schema, your cognitive machinery would be able to conclude that there is a self and there is an apple, but the construct of awareness would be missing.

People attribute awareness to other beings. Of course we attribute far more than the mental state of awareness. We routinely attribute emotions, motivations, intentions, and so on. But at the root of these attributions is the attribution of awareness. It is difficult to attribute to Jack a fear of a snake, or a motivation to run from the snake, or an intention to capture the snake, without also attributing to Jack an awareness of the snake. In effect, we attribute to Jack the property of having a mind that has encompassed and is focusing some of its resources on an item. The awareness we attribute to Jack serves as a schematic model of Jack's attentional process.

It is difficult to guess when in evolution an attention schema might have become adapted for social attribution. Figure 1 shows one possibility in which the social attribution of awareness became greatly expanded with the evolution of primates about 65 million years ago. Alternatively, it may of course have developed much earlier. Cats, dogs, even nonmammals such as birds might have the capacity to attribute awareness to others. This particular aspect of social attribution has not been systematically studied across phylogenetic lines.

In the human brain, an overlap may exist between the circuitry that attributes awareness to oneself and the circuitry that attributes awareness to others. Certain regions of the cortex are typically recruited during social perception as people construct models of other people's minds (e.g., Ciaramidaro et al., 2007; Saxe & Wexler, 2005; Saxe & Kanwisher, 2003; Vogeley et al., 2001; Brunet, Sarfati, Hardy-Baylé, & Decety, 2000; Gallagher et al., 2000; Fletcher et al., 1995; Goel, Grafman, Sadato, & Hallett, 1995). These regions include, among other areas, the STS, the TPJ, and the medial pFC. The medial pFC tends to be more active when people think about their own motivations and intentions and therefore, according to some speculations, may contribute to the construction of some of our psychological self-knowledge. But psychological self-knowledge is not the same thing as awareness. Self-knowledge is one domain of information about which a person can be aware. What about awareness of color, of temperature, of concept, of mathematics? What about awareness in general?

The TPJ has a pattern of response that may be particularly relevant to the question of awareness or at least to attributing awareness to someone else. The TPJ, bilaterally but with an emphasis on the right side, has been implicated in reconstructing the beliefs of others (Saxe & Wexler, 2005; Saxe & Kanwisher, 2003). For example, the TPJ becomes active when participants read a story about Sally and attribute to her the belief that her sandwich is in basket A instead of basket B. Although the term "belief" is commonly used, it is perhaps a misnomer. A better description of these experimental manipulations might be that Sally has certain information active in her mind. She is aware that the sandwich is in basket A. Her mind currently possesses that information. When participants are asked to consider what is currently in someone else's mind, answering that question reliably activates the TPJ. The critical property here is not the specific content—whether sandwiches, baskets, motivations, colors, or ideas—but instead whether that content is in someone else's mind or not.

In a contrasting line of research, it has been suggested that the TPJ might not be primarily involved in theory of mind, but instead may serve a more general role in attention. The TPJ, posterior STS, and ventral pFC are active in association with changes in one's own attentional state, especially when a novel or unexpected stimulus draws attention (e.g., Shulman et al., 2010; Mitchell, 2008; Astafiev, Shulman, & Corbetta, 2006; Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000). For this reason, these brain regions may be part of what has been termed the ventral attention network. Moreover, damage to the TPJ and STS can cause severe and long-lasting hemispatial neglect (Karnath, Ferber, & Himmelbach, 2001; Vallar & Perani, 1986). Indeed it appears that the most severe cases of hemispatial neglect occur with damage to the TPJ and not, as classically thought, to the parietal lobe.

Why should an area be involved in social cognition in some experiments and in attention in other experiments? One possible way to reconcile these two lines of research is that the regions of the brain that attribute awareness to other people in the context of social perception may also be necessary to attribute awareness to oneself. In this perspective, human awareness is something like color. It is a perception-like property computed by a specialized system in the brain. The computed property can be attributed to things. But unlike color, it must be attributed to two items, not one. Color is attributed to—or projected onto—a surface. Awareness is attributed to—or projected onto—a subject and an object. Agent Y is aware of thing X. I am aware of myself; I am aware of the blueness of the sky; the person next to me is aware of me; Alice is aware of my idea; John is aware of the stain on his shirt; whatever the specific references are, whoever is aware and whatever is the object of awareness, the attribution itself depends on similar underlying circuitry. That circuitry is active when we attribute awareness to others, and damage to that circuitry results in a loss of our own awareness of the things around us. That, at least, is the proposal.

Reprint requests should be sent to Michael S. A. Graziano, Department of Psychology, Princeton University, Princeton, NJ 08544, or via e-mail: graziano@princeton.edu.

REFERENCES

Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2006). Visuospatial reorienting signals in the human temporo-parietal junction are independent of response selection. *European Journal of Neuroscience, 23,* 591–596.

Baars, B. J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. J. Davidson,

G. E. Schwartz, and D. Shapiro (Eds), *Consciousness and Self Regulation* (p. 41). New York: Plenum Press.
Barlow, R. B., Jr., & Fraioli, A. J. (1978). Inhibition in the Limulus lateral eye in situ. *Journal of General Physiology, 71,* 699–720.
Beck, D. M., & Kastner, S. (2009). Top–down and bottom–up mechanisms in biasing competition in the human brain. *Vision Research, 49,* 1154–1165.
Bode, H. R., Heimfeld, S., Koizumi, O., Littlefield, C. L., & Yaross, M. S. (1988). Maintenance and regeneration of the nerve net in hydra. *American Zoology, 28,* 1053–1063.
Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature, 391,* 756.
Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage, 11,* 157–166.
Budd, G. E. (2008). The earliest fossil record of the animals and its significance. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 363,* 1425–1434.
Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2,* 200–219.
Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia, 45,* 3105–3113.
Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience, 3,* 292–297.
Crick, F., & Koch, C. (1990). Toward a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2,* 263–275.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.
Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Science, 5,* 16–25.
Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., et al. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57,* 109–128.
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia, 38,* 11–21.
Gazzaniga, M. S. (1970). *The bisected brain*. New York: Appleton Century Crofts.
Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport, 6,* 1741–1746.
Graziano, M. S. A. (2013). *Consciousness and the social brain*. New York: Oxford University Press.
Graziano, M. S. A., & Botvinick, M. M. (2002). How the brain represents the body: Insights from neurophysiology and psychology. In W. Prinz and B. Hommel (Eds), *Common mechanism in perception and action: Attention and performance XIX* (pp. 136–157). Oxford England: Oxford University Press.
Graziano, M. S. A., & Kastner, S. (2011). Human consciousness and its relationship to social neuroscience: A novel hypothesis. *Cognitive Neuroscience, 2,* 98–113.
Hwang, E. J., & Shadmehr, R. (2005). Internal models of limb dynamics and the encoding of limb state. *Journal of Neural Engineering, 2,* S266–S278.
Karnath, H. O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature, 411,* 950–953.
Kentridge, R. W., Heywood, C. A., & Weiskrantz, L. (2004). Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia, 42,* 831–835.
Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11,* 16–22.
Lackner, J. R. (1988). Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain, 111,* 281–297.
Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks, 17,* 861–872.
Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain, 106,* 623–642.
Mitchell, L. P. (2008). Activity in the right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18,* 262–271.
Mysore, S. P., & Knudsen, E. I. (2013). A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection. *Nature Neuroscience, 16,* 473–478.
Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Science, 13,* 416–424.
Newman, J., & Baars, B. J. (1993). A neural attentional model for access to consciousness: A global workspace perspective. *Concepts in Neuroscience, 4,* 255–290.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know—Verbal reports on mental processes. *Psychological Review, 84,* 231–259.
O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral Brain Science, 24,* 939–973.
Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences, U.S.A., 91,* 7398–7403.
Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *Neuroimage, 19,* 1835–1842.
Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43,* 1391–1399.
Schiff, N. D. (2008). Central thalamic contributions to arousal regulation and neurological disorders of consciousness. *Annals of the New York Academy of Sciences, 1129,* 105–118.
Shulman, G. L., Pope, D. L., Astafiev, S. V., McAvoy, M. P., Snyder, A. Z., & Corbetta, M. (2010). Right hemisphere dominance during spatial selective attention and target detection occurs outside the dorsal frontoparietal network. *The Journal of Neuroscience, 30,* 3640–3651.
Tallon-Baudry, C. (2011). On the neural mechanisms subserving consciousness and attention. *Frontiers in Psychology, 2,* 397.
Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin, 215,* 216–242.
Vallar, G., & Perani, D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia, 24,* 609–622.
van Swinderen, B. (2012). Competing visual flicker reveals attention-like rivalry in the fly brain. *Frontiers in Integrative Neuroscience, 6,* 96.
Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., et al. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage, 14,* 170–181.
Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269,* 1880–1882.