Automatic Brains—Interpretive Minds

Matthew Roser and Michael S. Gazzaniga

Dartmouth College

ABSTRACT—*The involvement of specific brain areas in carrying out specific tasks has been increasingly well documented over the past decade. Many of these processes are highly automatic and take place outside of conscious awareness. Conscious experience, however, seems unitary and must involve integration between distributed processes. This article presents the argument that this integration occurs in a constructive and interpretive manner and that increasingly complex representations emerge from the integration of modular processes. At the highest levels of consciousness, a personal narrative is constructed. This narrative makes sense of the brain's own behavior and may underlie the sense of a unitary self. The challenge for the future is to identify the relationships between patterns of brain activity and conscious awareness and to delineate the neural mechanisms whereby the underlying distributed processes interact.*

KEYWORDS—*neural correlates of consciousness; interpreter; integration*

Although it has been known for more than a century that particular parts of the brain are important for particular functions, the past decade of functional magnetic resonance imaging (fMRI) research has lead to a huge upsurge in evidence for functional specialization. This work has identified areas of the cortex, the convoluted outer layer of the brain, that are involved in processing particular stimulus attributes, or performing certain tasks. For example, cortical areas especially responsive to faces, movement, and places have been found, and these experimental results have been replicated by many independent observers. Although some of the initial claims for functional specialization have been tempered somewhat in the light of new findings, it is becoming ever more clear that the cortex is not a homogeneous, general-purpose computing device, but rather is a complex of circumscribed, modular processes occupying distinct locations.

Most of the work undertaken by these specialist systems occurs automatically and outside of conscious control. For instance, if certain stimuli trick your visual system into constructing an illusion, knowing that you have been tricked does not mean that the illusion disappears. The part of the visual system that produces the illusion is impervious to correction based on such knowledge. Additionally, a convincing illusion can leave behavior unaffected, as when observers are asked to scale the distance between their fingers to the size of a line presented with an arrowhead attached to each end. Although the arrowheads can alter the perceived size of a line (the Müller-Lyer illusion), observers do not make a corresponding adjustment in the distance between their fingers, suggesting that the processes determining the overt behavior are isolated from those underlying the perception. Thus, a visuo-motor process in response to a stimulus can proceed independently of the simultaneous perception of that stimulus (Aglioti, DeSouza, & Goodale, 1995).

Stimuli that are not consciously perceived by subjects can, nonetheless, affect behavior. For example, stimuli that are presented very briefly and followed by a masking stimulus go unperceived by subjects, but still activate response mechanisms and speed the recognition of following stimuli that share their semantic properties (Dehaene et al., 1998). When you add to this the observation that robust perceptual aftereffects can be induced by stimuli that are not consciously perceived (Rees, Kreiman, & Koch, 2002), it becomes evident that a great deal of the brain's work occurs outside of conscious awareness and control. Thus, the systems built into our brains carry out their jobs automatically when presented with stimuli within their domain, often without our knowledge.

The most striking evidence for the isolation of function from consciousness comes from studies of patients showing either neglect or blindsight. Neglect is a condition in which the patient ignores a part of space, usually the left; it is typically found in people with damage to the right parietal area of the brain and is thought to be due to the disruption of the brain's mechanisms for allocating attention. Astonishingly, patients with neglect often deny that they have any such condition. It is as if their consciousness of the deficit is destroyed by the lesion just as their actual awareness of a part of space is, even though early visual areas of the brain (i.e., areas that receive and process incoming visual information) are intact and functioning.

The even more bizarre condition known as blindsight describes the residual visual function shown by some patients following a lesion in early visual areas. Although these patients claim to be completely blind in the side of visual space contralateral to the lesion, they are nonetheless able to discriminate, locate, and guide motion toward a stimulus in that area, all without a conscious percept (Rees et al., 2002).

Together, these syndromes and studies in normal subjects suggest that the activity of the brain is not strictly continuous with our conscious experience. Instead, we are sometimes oblivious to complex

Address correspondence to Matthew Roser, Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hanover, NH 03755; e-mail: matthew.roser@dartmouth.edu.

processing that occurs in the brain. The question then becomes, what determines whether a process is conscious or not?

BRAIN ACTIVITY AND CONSCIOUSNESS

The neural correlates of consciousness in the human brain have been investigated using fMRI and a technique known as binocular rivalry. In this kind of study, a different stimulus is presented to each eye, and the conscious percept typically switches back and forth between the two stimuli, each percept lasting for a few seconds. Subjects indicate when their perception changes from one stimulus to the other, and because the stimuli themselves are static, any changes in neural activity that correlate with a change in the reported percept can be ascribed to changes in the contents of awareness (Tong, Nakayama, Vaughan, & Kanwisher, 1998). Brain activations elicited by rivalrous stimuli are very similar in magnitude and location to activations seen in response to separate stimuli that are presented alternately, suggesting that areas involved in processing a type of stimulus are also involved in the conscious perception of that type of stimulus (Zeki, 2003).

fMRI studies have also revealed substantial brain activations in response to stimuli that are not consciously perceived by subjects (Moutoussis & Zeki, 2002). For example, when color-reversed faces (e.g., an outlined red face on a green background and an outlined green face on a red background) are displayed separately to the two eyes, binocular fusion occurs, and subjects report seeing only the color that results from the combination of the two stimulus color (in this case, yellow). The color inputs to the brain are "mixed," like paint on an artist's palette, and the face stimuli become invisible. Despite not being consciously seen, these stimuli typically activate those areas of the brain that are activated by perceived faces. Why then are some seen and some not?

Brain activations correlated with perceived stimuli and those correlated with unseen stimuli show differences in both their intensity and their spatial extent (Dehaene et al., 2001). Dehaene and his colleagues found that although unperceived stimuli and perceived stimuli activated similar locations in the brain, the activations associated with perceived stimuli were many times more intense than those seen with unperceived stimuli and were accompanied by activity at additional sites. Thus, consciousness may have a graded relationship to brain activity, or a threshold may exist, above which activation reaches consciousness (Rees et al., 2002). At present this issue is unresolved, but the development of increasingly sophisticated designs in fMRI may yield progress by allowing the degree of activation to be determined as the availability of a stimulus to awareness is manipulated.

The increased spatial extent of activations elicited by perceived stimuli in the experiment by Dehaene and his colleagues suggests another possible mechanism for determining whether a stimulus reaches consciousness. Processing of a stimulus may reach consciousness if it is integrated into a large-scale system of cortical activity.

CONSCIOUSNESS SEEMS UNITARY

Despite the evidence that processing is distributed around the brain in functionally localized units, and that much processing proceeds outside of awareness, we personally experience consciousness as a unitary whole. How can these observations be resolved?

One possibility is that processes occurring within localized areas and circumscribed domains become available to consciousness only when they are integrated with other domains. Dehaene and Naccache (2001) have hypothesized that there is a *global neuronal work space* in which unconscious modular processes can be integrated in a common network of activation if they receive amplification by an attentional gating system. Attentional amplification leads to increased and prolonged activation and allows processing at one site to affect processing at another. In this way, brain areas involved in perception, action, and emotion can interact with each other and with circuits that can reinstate past states of this work space.

According to this hypothesis, consciousness is the collection of modular processes that are mobilized into a common neuronal work space and integrated in a dynamic fashion. It is a global pattern of activity across the brain, allowing information to be maintained and influence other processes. For instance, consider the task in which subjects are asked to match the distance between their fingers to the size of a Müller-Lyer figure. If a small delay is introduced between the observation of the figure and the reaching response, subjects must rely on their memory of the perceived size when scaling their grip to the size of the figure. Memory involves a consciously maintained representation. In this situation, the illusion does, in fact, affect the subject's motor response (Aglioti et al., 1995).

This model can explain some of the bizarre deficits of consciousness that occur as the result of brain lesions. As processing that does not achieve amplification remains entirely outside of consciousness, a neglect patient may not be aware of his or her deficit because the mechanism linking local processing to global patterns of activation has been disrupted. Thus, a lesion in a specific location may wipe out not merely processing of an attribute, but also the consciousness of the attribute.

Patients with severe cognitive deficits often confabulate wildly in order to produce an explanation of the world that is consistent with their conscious experience. These confabulations include completely denying the existence of a deficit and probably result from interpretations of incomplete information, or a reduced range of conscious experience (Cooney & Gazzaniga, 2003). Wild confabulations that seem untenable to most people, because of conscious access to information that contradicts them, probably seem completely normal to patients to whom only a subset of the elements of consciousness are available for integration.

MIND IS INTERPRETIVE AND CONSTRUCTIVE

The corpus callosum, which connects the two hemispheres, is the largest single fiber tract in the brain. What happens, then, when you cut this pathway for hemispheric communication and isolate the modular systems of the right hemisphere from those in the left? In the so-called split brain, only processes within a hemisphere can be integrated via cortical routes, and only a limited number of processes that can propagate via subcortical routes can be integrated between the hemispheres. Upon introspection, split-brain patients will tell you that they feel pretty normal. And yet, splitting the brain can reveal some of the most striking disconnections between brain processes and awareness. Each hemisphere can be presented with information that remains unknown to the opposite hemisphere.

Experimental designs that exploit this lack of communication have revealed that the left hemisphere tends to interpret what it sees, including the actions of the right hemisphere (Gazzaniga, 2000). For example, suppose two different scenes are presented simultaneously, one to each hemisphere, and the patient is asked to use his or her left hand to choose an appropriate item from an array of pictures of objects that may or may not be typically found within the presented scenes. The left hand is controlled by the right hemisphere, so the patient's left hemisphere, which has no knowledge of what was presented to the right hemisphere, can observe the subsequent actions of the right hemisphere. If the patient is asked why he or she chose a particular item, the patient's verbal reply will be largely controlled by the left hemisphere, where the brain's primary language centers are located. Studies using this procedure have shown that patients often reply with an interpretation of the action that is congruent with the scene presented to the left hemisphere. Thus, patients resolve one hemisphere's actions with the other hemisphere's perceptions, by producing an explanation that eliminates conflict between the two. Patients' responses in such studies are very similar to the confabulations produced by brain-damaged patients who deny that they have a serious deficit by rationalizing their bizarre behavior (Cooney & Gazzaniga, 2003).

The hypothesis-generating nature of the left hemisphere has also been demonstrated in a nonlinguistic manner. When each hemisphere of a split brain is asked to predict whether a light will appear on the top or the bottom of a computer screen on a series of trials, and to indicate its prediction by pushing one of two buttons with the contralateral hand, the two hemispheres employ radically different strategies. The right hemisphere takes the simple approach and consistently chooses the more probable alternative, thereby maximizing performance. By contrast, the left hemisphere does what neurologically normal subjects do and distributes its responses between the two alternatives according to the probability that each will occur, despite the fact that this is a suboptimal strategy (Wolford, Miller, & Gazzaniga, 2000). It seems that the left hemisphere is driven to hypothesize about the structure of the world even when this is detrimental to performance.

The left-hemisphere interpreter may be responsible for our feeling that our conscious experience is unified. Generation of explanations about our perceptions, memories, and actions, and the relationships among them, leads to the construction of a personal narrative that ties together elements of our conscious experience into a coherent whole. The constructive nature of our consciousness is not apparent to us. The action of an interpretive system becomes observable only when the system can be tricked into making obvious errors by forcing it to work with an impoverished set of inputs, such as in the split brain or in lesion patients. But even in the damaged brain, this system still lets us feel like "us."

CONCLUSIONS AND FUTURE DIRECTIONS

It is becoming increasingly clear that consciousness involves disunited processes that are integrated in a dynamic manner. It is assembled on the fly, as our brains respond to constantly changing inputs, calculate potential courses of action, and execute responses. But it is also constrained by the nature of modular processes that occur without conscious control, and large parts of it can be destroyed, leaving a rump that operates only within its reduced sphere. Progress toward an overarching theory of consciousness will involve putting our picture of the brain back together. Although carving cognition and brain function up at the joints has been a hugely productive approach, future progress must depend on a variety of approaches that integrate disparate and circumscribed processes.

To this end, developing techniques in brain mapping hold much promise. Statistical analysis of fMRI data allows the correlations between activations in different areas to be assessed, yielding maps of cerebral interactivity. The application of these techniques to investigation of the neural correlates of consciousness is extremely relevant, as the activation of large networks is thought to be a necessary condition for consciousness.

A further step involves integrating maps of cerebral interactivity with data about neuroanatomical connections. This technique allows a subset of brain processes to be explicitly modeled as a functional network and yields a map of the strengths of anatomical connections that best fits the imaging data (Horwitz, Tagamets, & McIntosh, 1999). At present, much of the data on neuroanatomical connections comes from postmortem studies in monkeys, but a developing noninvasive MRI technique known as diffusion-tensor imaging (DTI) allows the paths of neurons to be tracked and should provide more accurate data about the human brain. DTI is set to have a huge future impact on this field (Le Bihan et al., 2001).

The brain sciences of the coming years promise to yield great progress in our understanding of integrative processes in the brain. The ultimate aim is to come to a theory of consciousness that, while acknowledging that our brains are elaborate assemblies of myriad processes, explains how it is that we feel so unified.

Recommended Reading

Driver, J., & Mattingley, J.B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, *1*, 17–22.

Gazzaniga, M.S. (2000). (See References)

Savoy, R.L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica*, *107*, 9–42.

Acknowledgments—Preparation of this article was supported by Grant NS31443 from the National Institutes of Health. We are grateful to Margaret Funnell, Paul Corballis, and Michael Corballis for interesting discussions on the topics covered here.

REFERENCES

Aglioti, S., DeSouza, J.F., & Goodale, M.A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*, 679–685.

Cooney, J.W., & Gazzaniga, M. (2003). Neurological disorders and the structure of human consciousness. *Trends in Cognitive Sciences*, *7*, 161–165.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*, 1–37.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D.L., Mangin, J.F., Poline, J.B., & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*, 752–758.

Dehaene, S., Naccache, L., Le Clec, H.G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.

Gazzaniga, M.S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain*, *123*, 1293–1326.

Horwitz, B., Tagamets, M.-A., & McIntosh, A.R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences*, *3*, 91–98.

Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., & Chabriat, H. (2001). Diffusion tensor imaging: Concepts and applications. *Journal of Magnetic Resonance Imaging*, *13*, 534–546.

Moutoussis, K., & Zeki, S. (2002). The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences, USA*, *99*, 9527–9532.

Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, *3*, 261–270.

Tong, F., Nakayama, K., Vaughan, J.T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Wolford, G., Miller, M.B., & Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. *Journal of Neuroscience*, *20*, RC64.

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences*, *7*, 214–218.