

Is consciousness epiphenomenal? Social neuroscience and the case for interacting brains

John T. Cacioppo & Gary G. Berntson

Center for Cognitive and Social Neuroscience, 5848 S. University Avenue,
University of Chicago, Chicago, IL 60637
Cacioppo@uchicago.edu

Volume 3
Summer 2012

journal homepage
www.euresisjournal.org

Abstract

Consciousness has been cast as epiphenomenal. We show that this conclusion is based on the analysis of consciousness in the context of the individual human brain. The premise of social neuroscience is that the study of the human brain is incomplete when treated as a solitary organ. When the brain is viewed within a social context — that is, within a context of interacting brains — the material mechanism by which consciousness plays a role in subsequent brain states is revealed.

1. Introduction

Volition and consciousness have traditionally been important determinants in judgments of whether a person is culpable for an action. Volition, or free will, has been defined by philosophers as possessing the power to have done otherwise or to choose to do otherwise given the same circumstances. Free will, in turn, involves the notion that consciousness plays an instrumental role in what we feel, think, choose, and do. But what is the nature of consciousness? This is a question that has interested philosophers for centuries and one that scientists have now begun to ask. Rene Descartes' dualistic proposal that body and mind were distinct and separate had a profound effect on thought on the topic, including our notions of culpability. Dualism provided simple answers to these questions. Humans were cast as having a mechanical, determined, animalistic side and a conscious, volitional, spiritual side. Consciousness and free will were thought to be faculties of the latter, so the assignment of culpability required only that the offensive behavior not be a product of the reflexive, animalistic side of human nature.

Contrary to Descartes' approach, most scientific approaches to consciousness are grounded in monism and determinism. By monism, we mean that there is only one ultimate substance.

Monism does not mean that there is only one optimal *representation* of this substance, however. Different forms of representation make certain calculations easy and others difficult, as when using the periodic table to create new molecules versus using recipes to create dinner. In this context, monism means that the mind and consciousness are products of the operations of the central nervous system (CNS), which includes the brain and spinal cord but which for simplicity we will refer hereafter to as the brain.

By determinism we mean that there are specifiable conditions for everything that happens. Recent work in philosophy, psychology, psychiatry, and neuroscience has called into question the functional importance of consciousness and, with it, the notion of free will in human actions. The prediction of decisions based on brain functions observed prior to the decision has led to suggestions that the perception of free will is an illusion, that consciousness is epiphenomenal, and that people may not have as much responsibility for their actions as assumed. This notion is not entirely new, of course. As Hume observed, our actions are determined, in which case we are not responsible for them, or they are the result of random events, in which case we are not responsible for them.

The notion is that if consciousness is predetermined by antecedent brain functions, and if it can have no functional role in determining subsequent brain states, then consciousness would be without function — that is, consciousness would be epiphenomenal. We will return to this thesis, but we begin by distinguishing among the concepts of free will, consciousness, mind, and brain. We then examine key elements of the somato-motor machinery through which human behavior is expressed. We then outline *within a monist, deterministic framework* how the brain and consciousness may have reciprocal influences. If true, then consciousness is not epiphenomenal but rather has an important functional role in the operation of mind and brain. We conclude by considering some of the implications of such a state of affairs, including the notion of free will.

2. Definitions

The constructs of free will, consciousness, mind, and brain are like a set of Russian dolls, with each construct fitting within the domain of the next. We may *will* to do this or that, which is to say we exercise choice. The key element of *free will*, however, is that we could choose to do otherwise given the same circumstances. One can have the conscious experience of will and choice, and the conscious conviction that one could have chosen otherwise, but if consciousness is epiphenomenal then the feeling of *free will* is an illusion.

Hume reasoned that if it were an illusion, one could not be morally responsible for one's actions. There are reasons to reject this conclusion, whether or not free will is an illusion, however. Society requires that people be responsible for their actions. Even the behavior

of nonhuman primates is influenced by the reactions of others to their own actions. If an individual acts in a way that violates the rules of the group, the individual may be held responsible for these actions, and the rewards that are lost or the costs that are incurred can reduce the likelihood of repeat offenses against the collective. Importantly, this is the case whether or not consciousness is epiphenomenal and whether or not free will is an illusion. The same behavior enacted by accident is treated differently because the offensive behavior is less likely to re-occur even in the absence of social costs. In this way, social responsibility for one's deliberate actions is a fundamental building block of society.

By consciousness, we mean awareness, including the sensations, perceptions, emotions, thoughts, and mental images on which a person can report. Among the set of events that constitute our consciousness is the experience that we freely choose among various ways to act. I may write an essay that advocates a position contrary to my beliefs because I was instructed to do so as part of an experiment in which I was participating — in which case I would feel as if I had no choice, or because I chose freely to do so — in which case I would feel as if I could have chosen otherwise. Research on cognitive dissonance theory has shown that people can be randomly assigned to “low choice” or “high choice” conditions, and all can be induced to write the same counterattitudinal essay. That is, the experimenter determines whether a participant writes the essay with the conscious experience of having no other option but to do so, or the participant writes the essay with the conscious experience of having freely chosen to do so. In neither case was their actual choice — participants would not normally write such an essay without the experimenter's subtle manipulations of them. But whether or not they felt as if they had a conscious choice to do so greatly influenced what happened as a result of expressing the behavior. Participants in the low choice condition did not change their beliefs or behaviors toward the position they advocated in the essay because they felt no responsibility for this behavior. Participants in the high choice condition, however, changed their beliefs and attitudes to align them with the position they advocated in the essay because they consciously felt responsible for this behavior. This result has been interpreted to mean that conscious choice matters, even when it is an illusion (see review by [1]). But consciousness itself may be epiphenomenal, in which case conscious choice — or the (perceived) expression of free will — does not really matter, either — it would be the underlying brain states that matter.

Consciousness, then, includes the experience of free will but is not limited to free will. The mind, in contrast, refers to the structures and processes responsible for thought, emotion, and behavior. As such, the mind encompasses but extends beyond consciousness to include unconscious processes. For instance, Figure 1 illustrates the Stroop Test, a task that illustrates the influence of nonconscious processes on our thoughts and behaviors. The instruction to participants is simple: Identify the color of ink in which each letter string is printed. Words are then presented, one at a time, and participants announce the color in which the word is presented as quickly as they can. When the words are the names of colors (e.g., blue, red, or

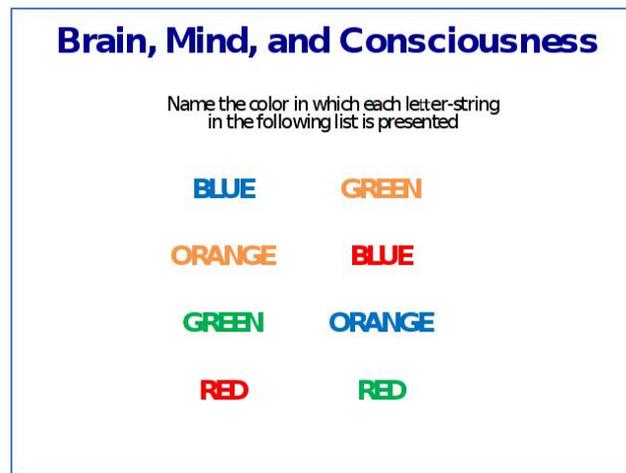


Figure 1: *The Stroop Test.*

ange) and are printed in the corresponding color (e.g., “blue” is printed in blue), participants perform the task quickly and with few if any errors (see Figure 1, left column). However, when the words are the names of colors printed in discrepant colors (e.g., “blue” is printed in red), participants stumble, taking longer to identify the colors and making more errors in identifying the colors in which the words are printed (see Figure 1, right column). The difference in behavior is because the brain is reading the words even though the participants are not consciously choosing to do so and, in fact, are not even consciously aware of doing so. When the words and colors are consistent, nonconsciously attending to the meaning of the words facilitates performance on the task. When the words and colors are inconsistent, however, this nonconscious information processing interferes with the conscious attempt to identify the color in which the words are printed. There are many such examples that could be given for preattentive or nonconscious processes of the mind, but the Stroop test is sufficient to illustrate that the mind includes conscious experience but its operations are not limited to conscious experiences [2].

Finally, the brain is the organ of the mind but also performs functions that are not directly relevant, including autonomic, neuroendocrine, and immune function such as the modulation of the regulatory mechanisms involved in the maintenance of homeostatic set-points for body temperature, blood pressure, and blood sugar. If the brain were to fail in these operations, there would be dramatic consequences for the mind, consciousness, and behavior but the same could be said about the failure of any major organ (e.g., the heart). Thus, free will (whether an illusion or not) represents an experience that falls within the larger domain of consciousness, consciousness (whether epiphenomenal or not) represents a set of structures and processes within the larger domain of the mind, which in turn represents a set of structures and processes within the larger domain of the brain.

3. The machine

A common misconception about human behavior is that it follows in a straightforward fashion from beliefs and intentions, enacted by the homunculus positioned in the primary motor cortex. Dennett 1991 [3] describes such an enactment as the Cartesian theater and regards it as a remnant of instinctively dualistic thinking. One problem with this conception is that it accounts for human agency and behavior in a circular fashion [4]. A second is that the primary motor cortex is only a small piece of a heterarchically organized motor system. A third is that the cell bodies of the motor system themselves represent but a small part of the complex information processing capacities of the brain (and CNS more generally).

Behavior is complexly determined. Neurobiologically, human action is enacted through the motor system, which controls somatic muscles of skeletal movement. The muscles are innervated by the lower and upper motoneuron systems. The lower motoneuron system, sometimes called the final common pathway, consists of cell bodies located in the spinal cord, brain stem, and pons with axon fibrils extending to motor end plates on muscle fibers. A lesion of any part of the lower motoneuron system results in a flaccid paralysis (loss of motor control and muscle tone) of the muscles to which the severed motoneurons had projected.

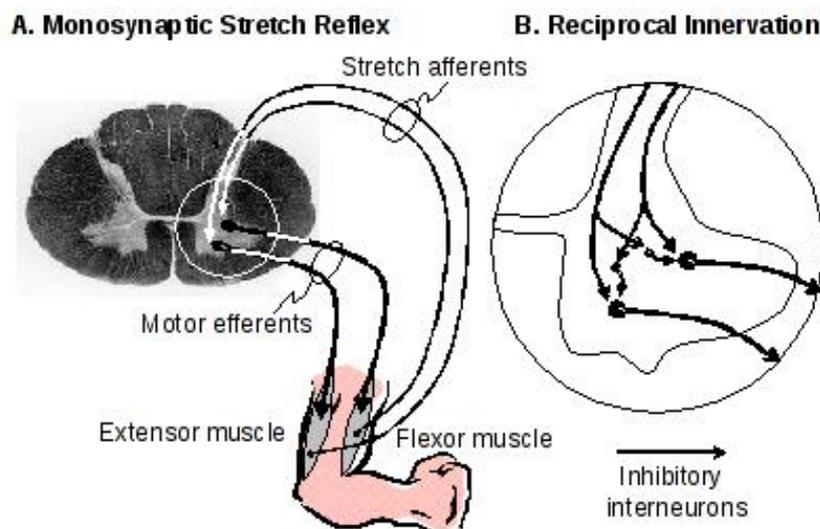


Figure 2: *Stretch Reflex. A. Basic parallel circuits of the flexor and extensor stretch reflex. B. Reciprocal innervation, an example of Sherrington's alliance of reflexes. Dotted lines represent inhibitory interneurons which achieve a level of reciprocal integration between flexor/extensor motor neuron pools. (From [7].).*

As documented by early investigators, such as [6], basic somatomotor control is effected at the level of the spinal cord, with spinal reflexes representing the lowest central level in

somatomotor control systems (Figure 2). Spinal reflex circuits are relatively simple and may be comprised of a single central synapse. The monosynaptic stretch reflex is exemplified by the knee jerk reflex to the physician's rubber mallet. This reflex entails an afferent somatosensory link arising from muscle stretch receptors, which synapse directly on the lower motor neurons controlling that muscle. This simple circuit provides for a reflexive contraction of the stretched muscle, which tends to compensate for the perturbing stretch. Stretch reflexes exist in all major classes of somatic muscles, including flexor and extensor muscles.

Opponent flexor (e.g., biceps) and extensor (e.g. triceps) reflexes antagonize one another and promote opposite outcomes for the limb (flexion and extension, respectively). The basic neural circuits of these reflexes are independent and organized in parallel, and they have limited inputs and outputs, allowing for rapid, efficient processing (Figure 2). The cost of this efficiency, however, is that lower-level systems have limited integrative capacity. Moreover, they can be in conflict. Simultaneous stretch of both the flexor and extensor muscle may lead to a reflexive increase in muscle tension in both muscles, but because they are opposed in their actions, there may be no resultant limb movement.

Greater levels of integration in motor systems are achieved by hierarchical circuits that promote coordination among the basic spinal reflexes – what Sherrington referred to as the *alliance of reflexes*. In our flexor/extensor example, this entails a collateral projection of the stretch receptor afferents onto inhibitory interneuron circuit elements, which in turn project to and inhibit the motor neuron for the opposing muscle. Stretching the flexor muscle, for example, results not only in activation of the flexor motor neurons (stretch reflex), but also inhibition of the opposing motor neurons via an inhibitory neural pathway. This exemplifies a general principle of neural organization articulated by Sherrington — the principle of *reciprocal innervation* — which stipulates that neural systems promote specific outcomes by activating the mechanisms for the target response while at the same time inhibiting opposing responses [7, 8].

Sherrington's alliance of reflexes does not stop with reciprocal innervation. The cell bodies of the lower motoneuron system are innervated by upper motoneurons from the pyramidal and the extrapyramidal systems. A lesion of each of these upper motoneuron systems has distinctive behavioral effects. The cell bodies of the pyramidal upper motoneuron system are located in the precentral gyrus of the frontal lobe, and long axons project to the cell bodies of the lower motoneuron system in the pons, brain stem, and spinal cord. The pyramidal motor system controls all our voluntary movements — so called because movements enacted through the pyramidal system tend to be associated with the conscious experience of having chosen to make the movement, even when this conscious experience trails measurable brain activities predicting the movement and precursors of the movement itself by seconds (e.g., [9]). For instance, a lesion of the pyramidal tract from the facial region of the motor cortex to the cell bodies in the facial nerve nucleus results in the inability to volitionally smile on

the contralateral side of the face. If the afflicted individual hears a humorous joke, however, a normal smile is produced [34].

The ability exists to produce a symmetrical smile spontaneously/reflexively but not volitionally because the extrapyramidal upper motoneuron system is still intact. The extrapyramidal upper motoneuron system is phylogenetically older than the pyramidal system and thus plays a relatively more important role in lower animals. The nuclei constituting the extrapyramidal system are manifold with the major parts located in the basal ganglia, including the caudate, putamen, and globus pallidus; the thalamus and subthalamic nucleus; and the substantia nigra and red nucleus in the midbrain. All of these nuclei are connected synaptically to one another, the brain stem, cerebellum, and the pyramidal system. A lesion of the extrapyramidal innervation of the facial nerve nucleus leaves the afflicted individual capable of producing a voluntary symmetrical smile, but the individual is incapable of producing a normal spontaneous smile when they hear something humorous [34].

As hierarchical levels are layered on the motor control system, progressively higher levels receive a wider array of inputs, have greater circuit complexity and computation capacity, and can achieve a broader and more flexible range outputs [7, 8]. At the highest levels, beyond the primary motor cortex, cerebral systems must process a tremendous amount of sensory information, and integrate this information with associative networks, emotional/motivational substrates, and expectancies, in the contexts of strategic goals and tactical plans. This requirement for enhanced information processing can impose a processing bottleneck that necessitates a slower, more serial mode of processing and selective attentional mechanisms, e.g. [10]. Although it is these highest level systems that confer the greatest cognitive and behavioral capacity, they do not operate in isolation but depend upon and interact with lower levels in the hierarchy.

Hierarchical dimensions of central nervous system organization can be demonstrated anatomically as well as functionally — see [5]. The simple hierarchy depicted by the solid lines in Figure 3, however, belies the true complexity of neurobehavioral substrates, as long ascending and descending pathways (dashed lines) can bypass intermediate levels of hierarchical organization and interconnect across widely separated neural levels. Cortical motor neurons project not only to intermediate-level somatomotor networks, but also directly onto spinal motor neurons through long descending pathways [11, 18]. The long ascending and descending pathways in neural hierarchies, together with the existence of lateral interactions among elements (such as those that underlie reciprocal innervation), yield what has been termed a heterarchical organization — see [5]. The outputs of a strict hierarchical system are coherent, as all levels are linked by intermediate regulatory levels, and all outputs are by final common pathways. In a heterarchical organization, however, higher levels can directly access output mechanisms independent of intermediate levels. This organizational feature allows for concurrent expression of multiple re-representative systems, which can increase

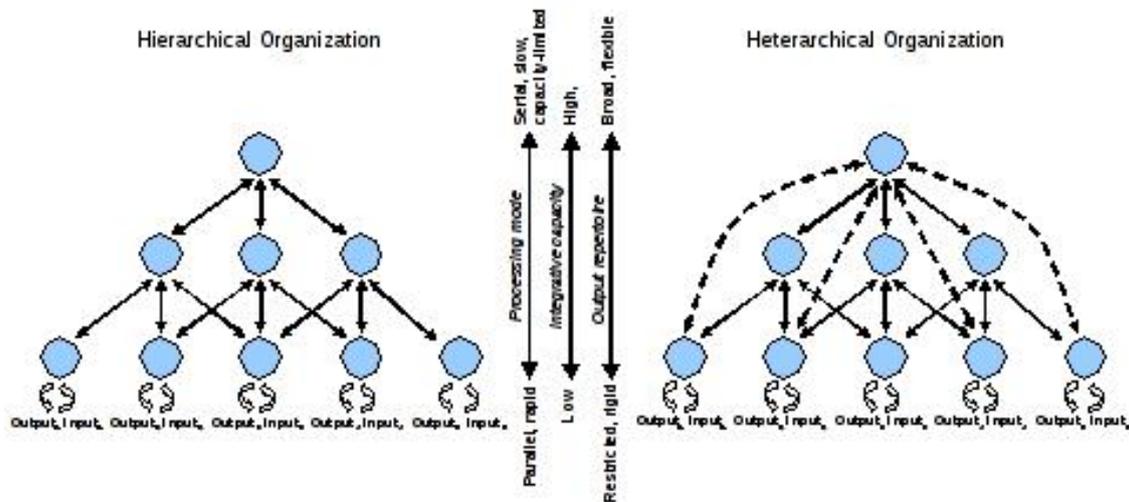


Figure 3: *Hierarchical and heterarchical organizations. A heterarchy differs from a hierarchy in the existence of long ascending and descending pathways that span intermediate levels. Properties of the levels in both classes of organizations lie along the illustrated continua of processing mode, integrative capacity and output repertoire. Heterarchical organizations, however, have greater integrative capacity and output flexibility as the long ascending and descending projections provide inputs and outputs that are not constrained by intermediate levels. (From [7].).*

behavioral complexity but can also lead to functional conflicts. By volitionally stiffening the leg, for example, higher level systems can mask the stretch reflex and usurp control of motor neuron pools.

Paralleling the pattern for somatic motor control, adaptive and protective reflexes and mechanisms are organized at all levels of the neuraxis. The pain withdrawal reflex, for example, is organized at the level of the spinal cord and can be seen even after spinal transections that isolate the cord from higher brain systems. Pain withdrawal reflexes are protective reactions that arise from somatosensory afferents, carrying nociceptive signals, which anatomically link to flexor neuron pools by a multisynaptic spinal pathway. Through this spinal reflex circuit, noxious stimuli yield a protective flexor withdrawal response. Likely because of their adaptive value, pain withdrawal reflexes are among the earliest to develop and the most resistant to disruption. Although pain withdrawal reflexes at the level of the cord may not require the invocation of a construct of emotion or affect, they represent an important low level evaluative mechanism for escape from noxious stimuli. Moreover, despite their neural simplicity, these circuits can show operant conditioning of escape which can support not only escape from, but active avoidance of, pain stimuli [4].

In contrast to the primitive avoidance system associated with flexor reflexes, separate mechanisms exist at the level of the cord for opposing extensor responses which promote engagement with the environmental stimuli. Reflexes such as the extensor reflex in response to non-painful cutaneous stimulation of the palm or the sole of the foot contribute to postural,

locomotory and grasping responses that serve to engage the organism with the environment.

As considered above, flexor and extensor reflexes are organized largely in parallel, as they control distinct motoneuron pools for opposing muscles. Nevertheless, they do interact. These lower reflex substrates are integrated by higher-level circuits, such as those that implement reciprocal innervation, which tends to reduce concurrent activation. They are also impacted by even higher systems that contribute to volitional actions and confer a greater degree of flexibility and control over flexor and extensor motor neuron pools and lower reflex substrates. Thus, we can volitionally contract both flexor and extensor muscles (e.g., in stiffening the arm) which can overcome the lower level reciprocal innervation, and we are able to override or suppress flexor pain withdrawal reflexes (e.g., to remove a sliver from the finger).

Given there are multiple levels of neurobehavioral control within the CNS and the responses these nuclei activate can differ in a given circumstance (e.g., as when the point of a needle containing a life-saving vaccine punctures the skin), response conflicts can occur. The frontal regions have long been thought to be involved in executive functions such as dealing with response conflicts, formulating goals and plans, selecting among options to achieve these goals, monitoring the consequences of our actions in light of our goals, and inhibiting, switching and regulating our behaviors accordingly. Aron, 2008 [12] reviews evidence that the initiation of a motor response proceeds from the planning areas of the frontal cortex to the putamen, globus pallidus, thalamus, primary motor cortex, motor nucleus in the spinal cord, and finally to the muscles. Being able to inhibit a motor response once it has been initiated has obvious adaptive value, and this inhibition involves the right inferior frontal cortex, which projects to the subthalamic nucleus (a region of the basal ganglia that may act on the globus pallidus to block the motor response). Monitoring for response conflicts, in turn, appears to involve the dorsal anterior cingulate and the adjacent presupplementary motor area which, in turn, is connected to the right inferior frontal cortex and subthalamic nucleus. Switching also involves the presupplementary motor area and right inferior frontal cortex. Together, this work has led to a model in which the presupplementary motor area may monitor for conflict between an intended response and a countervailing signal and when such conflict is detected the “brakes” could be put on via the connection between right inferior frontal cortex and the subthalamic nucleus region [12]. The term “monitoring” can be misleading, however. There is no homunculus overseeing our actions, but rather in a quite deterministic fashion the activation of multiple, sometimes conflicting, response circuits that include excitatory and inhibitory connections leads to the activation of some responses over others based on which first reaches its response threshold.

There is also an extensive behavioral literature showing that behavioral intentions predict volitional behaviors, and that the experimental manipulation of these behavioral intentions leads to corresponding changes in volitional behavior. In the theory of planned behavior

[13], intentions are determined by three distinct constructs: attitude toward the behavior, subjective norms about the behavior, and perceived behavioral control. Attitude, in turn, is determined by *beliefs* (b_i) and *evaluations* (e_i) of those beliefs such that the $attitude = \sum_i b_i e_i$ for $i = 1, n$. Subjective norm is determined by *normative beliefs* (nb_j) and the *motivation to comply* with these beliefs (mc_j) such that the $subjective\ norm = \sum_j nb_j mc_j$ for $j = 1, m$. And finally, the perceived behavioral control is determined by *control beliefs* (cb_k) and the *perceived probability* that these beliefs are true (pp_k) such that $perceived\ behavioral\ control = \sum_k (cb_k pp_k)$ for $k = 1, o$.

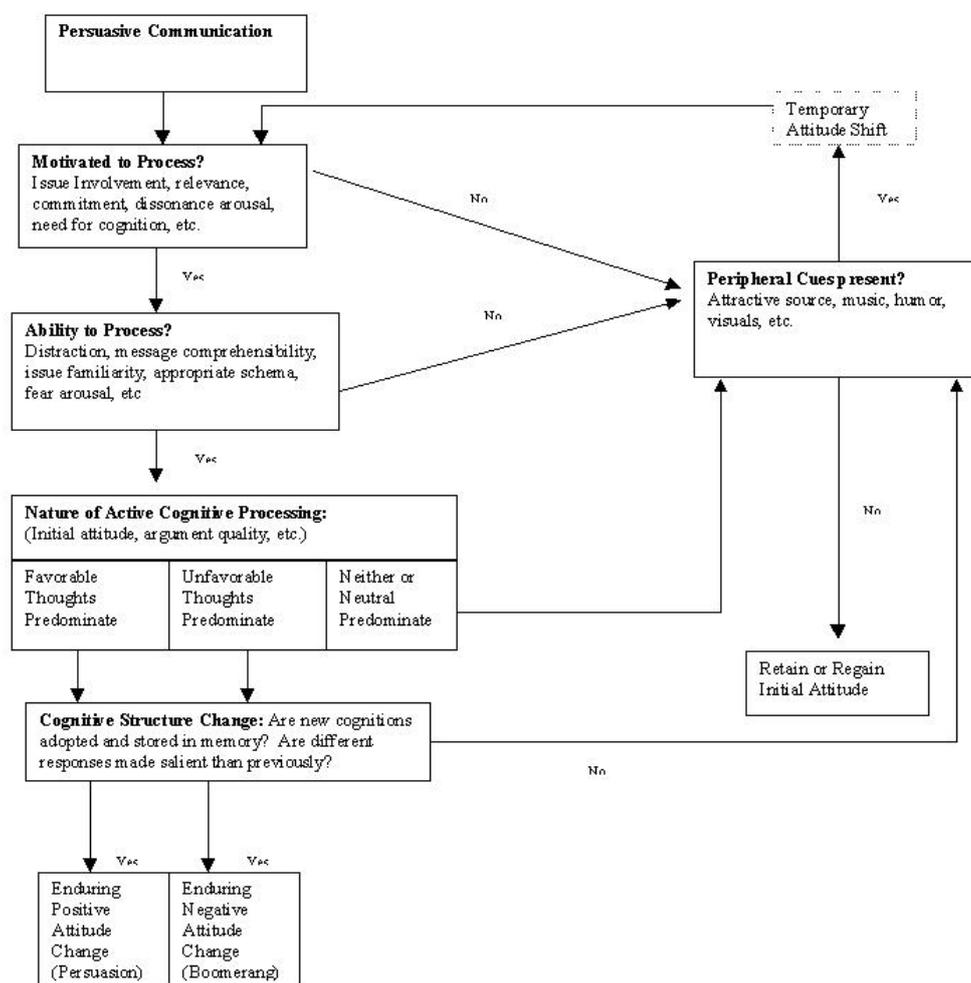


Figure 4: *The Elaboration Likelihood Model. (From [14].)*

Where do the beliefs and evaluations of beliefs that lead to the attitude come from? Hovland and colleagues suggested that attitudes were the result of comprehending and encoding the message arguments (i.e., belief statements). This view proved to be incorrect, and the message learning approach was replaced by the elaboration likelihood model (ELM) of attitude



change ([1, 14]; Figure 4). The ELM specifies that there are both effortful (central route) and non-effortful (peripheral route) attitude change processes, that attitude change derived through the central route can reflect rational or psycho-rational issue-relevant thinking, and that the central and peripheral route to attitude change operate under different but specifiable conditions. Finally, according to the ELM, there are different consequences of thoughtful and non-thoughtful persuasion such as attitudes derived through the central route are more enduring, more resistant to counterpersuasion, and more predictive of behavior than are attitudes derived through the peripheral route. There is now an extensive behavioral literature supporting the predictions of the ELM. For instance, multiple arguments can produce thoughtful change under some conditions (e.g., under high personal relevance of the message people scrutinized the merits of the arguments), but can produce the same amount of unthoughtful change under other conditions (e.g., under low personal relevance of the message people based their attitudes simply on the number of arguments presented regardless of their merit). The former thoughtful attitude change, however, is more persistent and predictive of behavior than the latter non-thoughtful change.

In the past decade, theory and research has led to an appreciation for the importance of the role of automatic associative processes, or implicit attitudes, especially in the production of spontaneous behaviors, e.g. [30, 16]. These have been termed “implicit” because they do not involve consciousness, at least in the same way as beliefs, attitudes, intentions, and decisions. Implicit processes have traditionally been depicted as representing a single, unitary mechanism, but the heterarchical organization outlined here suggests these processes, though no less deterministic, are multifarious in underlying mechanism and function.

4. The conundrum

The work on implicit attitudes and behaviors expands rather than contradicts work on the experience of agency (i.e., the awareness of being in control both of one’s own actions and through them of events in the external world). The prediction of decisions based on brain functions observed prior to the decision have led to suggestions that people may not have as much responsibility for their actions as assumed, however. The argument for consciousness being epiphenomenal can be parsed into two parts. The first part is that consciousness is fully predicted and determined by the brain’s prior and ongoing states. The second part is equally important, however, and it is the notion that consciousness plays no role in subsequent brain states. From the perspective of the brain, it is difficult to conceive through what material mechanism consciousness could conceivably alter subsequent brain states above and beyond the prior and ongoing brain states that give rise to consciousness. This is what philosophers call overcausation. If there is a sufficient causal explanation in the physical domain, there cannot be a meaningful explanation in some other domain. Without the specification of such a mechanism, one is left to conclude that consciousness has no functional role and,

hence, is epiphenomenal.

The argument is outlined more formally in Equation 4.1. The quantity β_t represents the brain state at any given point in time, designated as t . This brain state is designated to be a function of a constant representing the unique capacities of a given individual’s brain, the observed brain state at time $t - 1$, and the prediction error by the brain at time $t - 1$, which represents the difference between the state of the world (and brain) predicted for time $t - 1$ and the state of the world (and brain) at time $t - 1$. Our brains are metabolically expensive. They represent approximately 5% of our body mass but are responsible for approximately 20% of our oxygen consumption. This prediction error can be viewed as operating like a compression algorithm to minimize these metabolic costs. To send a video signal digitally, for instance, the information in each pixel of each frame of a film can be transmitted, but this is resource intensive and expensive. To make this transmission more efficient, one can send each pixel of the first frame and the pixels that have changed in each succeeding frame. Using such a compression algorithm greatly reduces the transmission load and costs. Of course, to know which pixels have changed from frame to frame, one has to compare what is predicted at each pixel based on the prior frame and what is observed. If there is a difference, then the new state of the pixel is transmitted and the image is updated. The use of such compression algorithms provides significant savings in terms of processing load to achieve transmission of the same information. Dating back to the classic work of Rescorla and Wagner (1968) [33], learning theorists have emphasized prediction error as an important component of updating our cognitive representations of the world and improving our response repertoires.

$$\begin{aligned} \text{Given } \beta_t &= f(\mu + \alpha_{i,t-1}\beta_{t-1}^{obs} + \alpha_{j,t-1}(\beta_{t-1}^{obs} - \beta_{t-1}^{pred}) + \dots), \\ \text{Determinism: } &\beta_t \rightarrow \beta_{t+1} \rightarrow \beta_{t+2} \rightarrow \beta_{t+3} \dots \\ \text{Monism: } &\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\ &\chi_t \quad \chi_{t+1} \quad \chi_{t+2} \quad \chi_{t+3} \dots \end{aligned} \tag{4.1}$$

- Cases to consider:
 1. walking along a smooth sidewalk;
 2. walking along a smooth sidewalk when you surprisingly step into an unforeseen depression, leading you to verbally express an expletive (χ_{t+1}).
- Importantly, in no instance does the portion of consciousness, χ , even when acute and expressed as an expletive have any influence, directly or indirectly, on subsequent brain states, leading to the conclusion that consciousness is epiphenomenal.

As also depicted in Equation 4.1, brain state at time t predicts, and in fact determines, the brain state at time $t + 1$, and so on. This follows from determinism, a basic neuroscientific tenet. Determinism has been challenged, for instance, based on arguments that quantum mechanics have shown events are not deterministic and, therefore, brain states cannot be

fully determined by prior brain states. Pinker (2003) [18] has noted that the neural operations of the brain are at a sufficiently macroscopic level that the uncertainty of quantum physics does not translate into the same uncertainty at the level of neuronal activation. With approximately 100 billion neurons and an estimated 10^{14} neuronal connections in the brain, we are far from being able to specify the precise mechanism through which each brain state determines the succeeding. However, scientific determinism does not mean one has a complete explanation, only that such an explanation exists.

The bottom line in Equation 4.1 illustrates how consciousness at time t , depicted as χ_t , is a function of the brain state at time t . Consciousness, in short, is a product of the functions of the brain -- consistent with the principle of monism. As such, however, it is left with no more function than the output display on a computer screen, playing no role in determining the succeeding state of the computer that produced this output.

Consider the case of an individual walking along a straight, smooth path. Each step is orchestrated by the brain and is a close replica of the preceding step. There is nominal prediction error, so β_t determines β_{t+1} which determines β_{t+2} and so forth. The corresponding state of consciousness at each point in time can be expressed as χ_t , χ_{t+1} , χ_{t+2} , and so on, but there is no known mechanism through which consciousness can influence the succeeding brain state. Although counterintuitive and perhaps implicitly objectionable, this line of reasoning leads to the conclusion that consciousness is epiphenomenal.

Now consider the case of an individual walking along a straight, smooth path who is surprised when a step falls into unnoticed depression along the path, leading the person at time $t + 1$ to express of what they suddenly became aware, "now that was a surprise," represented by χ_{t+1} . The difference between this case and the prior case is that there is a significant prediction error at time $t + 1$, and with this prediction error comes a heightened sense of awareness of some of the events that the brain was processing. Despite these differences in brain states and corresponding conscious states, the prediction error is included in the definition of β_t so, as in the prior case, β_t determines β_{t+1} which determines β_{t+2} and so forth, and the corresponding state of consciousness at each point in time, expressed as χ_t , χ_{t+1} , χ_{t+2} , and so on, is determined by the corresponding brain state. In this case, χ_{t+1} is associated with the statement, "now that was a surprise," but this difference is of no real consequence. There still is no known mechanism through which consciousness or this expression could influence the succeeding brain state. Again, then, this line of reasoning leads to the conclusion that consciousness serves no function.

5. The solitary brain versus the social brain

The analyses of consciousness to date have treated the solitary human brain as the appropriate unit of analysis. However, we are a social species, and social species, by definition, form organizations that extend beyond the individual. These structures evolved hand in hand with behavioral, neural, hormonal, cellular, and genetic mechanisms to support them because the consequent social behaviors helped these organisms survive, reproduce, and care for offspring sufficiently long that they too reproduced, thereby ensuring their genetic legacy. Social neuroscience is the interdisciplinary academic field devoted to understanding how biological systems implement social processes and behavior, and how these social structures and processes impact the brain and biology. Social neuroscience is not a cognitive neuroscience of social stimuli.

Behavioral neuroscience is a perspective in which the nervous system and brain are viewed as instruments of sensation and response. Research representing this perspective tends to focus on topics such as learning, memory, motivation, homeostasis, sleep and biological rhythms, and reproduction — and on the neural mechanisms underlying these behavioral functions. Cognitive neuroscience emerged as a distinct functional perspective in which the brain is viewed as an information processing organ, with a focus on topics such as attention, perception, representations, decision-making, memory systems, heuristics, reasoning, and executive functioning — and on the neural mechanisms in the human brain that underlie these representations and processes [19]. Social neuroscience represents yet another broad perspective that extends beyond the structure and function of a single organism to investigate the functions that are altered by or are derived from the association or interaction of conspecifics (imagined or real)— and on the neural and hormonal mechanisms underlying these structures and functions [20, 26]. If cognitive neuroscience is equivalent to the study of a computer connected to an electrical outlet, social neuroscience is equivalent to the study of a mobile, broadband connected computer linked to countless others via the internet.

Human social processes were once thought to have been incidental to learning and cognition, whereas the social complexities and demands of primate species are now thought to have contributed to the evolution of the neocortex and various aspects of human cognition. According to Dunbar and colleagues, e.g. [21], deducing better ways to find food, avoid perils, and navigate territories has adaptive value for large mammals, but the complexities of these ecological demands are no match for the complexities of social living (especially in hostile between-group social environments), which include: recognizing ingroup and outgroup members; learning by social observation; recognizing the shifting status of friends and foes; anticipating and coordinating efforts between two or more individuals; using language to communicate, reason, teach, and deceive others; orchestrating relationships, ranging from pair bonds and families to friends, bands, and coalitions; navigating complex social hierarchies, social norms and cultural developments; subjugating self-interests to the interests of

the pair bond or social group in exchange for the possibility of long term benefits for oneself or one's group; recruiting support to sanction individuals who violate group norms; and doing all this across time frames that stretch from the distant past to multiple possible futures. Consistent with this reasoning, human toddlers and chimpanzees have similar cognitive skills for engaging the physical world but toddlers have more sophisticated cognitive skills than chimpanzees for engaging the social world [4]; cross-species comparisons have revealed that the evolution of large and metabolically expensive brains is more closely associated with social than ecological complexity [23]; and a composite index of sociality in troops of baboons has been found to be highly correlated with infant survival [6].

Our survival depends on our connection with others. Born to the most prolonged period of utter dependency of any animal, human infants must instantly engage their parents in protective behavior, and the parents must care enough about their offspring to nurture and protect them. If infants do not elicit nurturance and protection from caregivers, or if caregivers are not motivated to provide such care over an extended period of time, then the infants will perish along with the genetic legacy of the parents [24]. Our developmental dependency mirrors our evolutionary heritage. Hunter/gatherers did not have the benefit of natural weaponry, armor, strength, flight, stealth, or speed relative to many other species. Human survival depended on *collective* abilities, not on individual might. *Communication* is critical to organizing these collective abilities.

Which brings us to two additional cases to consider in our investigation of the function of consciousness. First, consider an individual walking along a straight, smooth path who *imagines* another individual surprisingly stepping out from behind an obstruction, leading the first person at time $t + 1$ to express of what they suddenly became aware, “now that was a surprise,” represented by χ_{t+1} (see Equation 4.2). As in the prior case, there is a significant prediction error at time $t + 1$, and with this prediction error comes a heightened sense of awareness of some of the events that the brain was processing. Despite these differences in brain states and corresponding conscious states, the prediction error is included in the definition of β_t so, as in the prior case, β_t determines β_{t+1} which determines β_{t+2} and so forth, and the corresponding state of consciousness at each point in time, expressed as χ_t , χ_{t+1} , χ_{t+2} , and so on, is determined by the corresponding brain state. The other individual's brain states and consciousness are also constructs of the first person's brain states — so designated by their enclosure in the thought-bubble — so in fact these do not represent the true brain or conscious states of another individual, but only the constructions of a solitary brain. Accordingly, χ_{t+1} is associated with the statement, “now that was a surprise,” but the expression again has no real consequence, and we are left to conclude that consciousness serves no function.

$$\begin{aligned}
 &\text{Given } \beta_t = f(\mu + \alpha_{i,t-1}\beta_{t-1}^{obs} + \alpha_{j,t-1}(\beta_{t-1}^{obs} - \beta_{t-1}^{pred}) + \dots), \\
 &\quad \text{Determinism: } \beta_t \rightarrow \beta_{t+1} \rightarrow \beta_{t+2} \rightarrow \beta_{t+3}\dots \\
 &\quad \text{Monism: } \begin{array}{cccc} \downarrow & \downarrow & \downarrow & \downarrow \\ \chi_t & \chi_{t+1} & \chi_{t+2} & \chi_{t+3} \dots \end{array} \\
 &\quad \text{Imaginary Process: } \left\{ \begin{array}{cccc} \chi_t & \chi_{t+1} & \chi_{t+2} & \chi_{t+3}\dots \\ \uparrow & \uparrow & \uparrow & \uparrow \\ \beta_t \rightarrow & \beta_{t+1} \rightarrow & \beta_{t+2} \rightarrow & \beta_{t+3} \dots \end{array} \right. \tag{4.2}
 \end{aligned}$$

- Cases to consider:
 1. walking along a smooth sidewalk when *you imagine* an unforeseen person surprisingly step out from behind an obstruction, leading you to say “that was a surprise” (χ_{t+1})

Finally, consider the case of an individual walking along a straight, smooth path when another individual surprisingly steps out from behind an obstruction, leading the first person at time $t + 1$ to express of what they suddenly became aware, “now that was a surprise,” represented by χ_{t+1} . The brain and conscious states of the second person (shown in Equation 4.3) follow the same principles as those of the first person, adhering to determinism and monism. As in the prior case, there is a significant prediction error at time $t + 1$ when the first person encounters the second, and with this prediction error comes a heightened sense of awareness of some of the events that the brain was processing. Unlike the prior cases, however, the verbal expression of the part or whole of the contents of χ_{t+1} , as “now that was a surprise” serves as a communication, thereby producing an effect on the succeeding brain state of the second person (depicted in Equation 4.3). The brain of the first person continues to predict the likely succeeding brain state, including in this case an attempt to predict the thoughts, feelings, and behavior of the second individual — efforts that have been referred to as “mentalizing” and “theory of mind.” Typically, the second person’s conscious and behavioral responses are not perfectly predictable, however. In this particular case, the second person may respond to the communication, “now that was a surprise” with an apology for surprising the individual, an insulting expletive, a request for directions, and so forth. This person’s response to the expression of the first individual’s conscious content therefore influences that individual’s subsequent brain state (depicted in Equation 4.3). In this instance, the portion of consciousness, x_{t+1} , expressed as surprise, has an indirect but deterministic influence on that individual’s subsequent brain states. This line of reasoning implies that consciousness can play a deterministic, monistic, and functional role in brain states.

$$\begin{aligned}
 &\text{Given } \beta_t = f(\mu + \alpha_{i,t-1}\beta_{t-1}^{obs} + \alpha_{j,t-1}(\beta_{t-1}^{obs} - \beta_{t-1}^{pred}) + \dots), \\
 &\text{Determinism: } \beta_t \rightarrow \beta_{t+1} \xrightarrow{\chi_{t+1}} \beta_{t+2} \rightarrow \beta_{t+3} \dots \\
 &\text{Monism: } \begin{matrix} \downarrow & \downarrow & \downarrow & \downarrow \\ \chi_t & \chi_{t+1} & \chi_{t+2} & \chi_{t+3} \dots \end{matrix} \\
 &\text{Second individual: } \left\{ \begin{matrix} \chi'_t & \chi'_{t+1} & \chi'_{t+2} & \chi'_{t+3} \dots \\ \uparrow & \uparrow & \uparrow & \uparrow \\ \beta'_t \xrightarrow{\chi_{t+1}} & \beta'_{t+1} \rightarrow & \beta'_{t+2} \rightarrow & \beta'_{t+3} \dots \end{matrix} \right. \tag{4.3}
 \end{aligned}$$

- Cases to consider:
 1. walking along a smooth sidewalk when an unforeseen person surprisingly steps out from behind an obstruction, leading you to say “that was a surprise” (χ_{t+1}), to which the person responds (unpredictably) by apologizing by surprising you, expresses an expletive, asking you directions, walking by without expressing anything, etc.
- Importantly, in this instance the portion of consciousness (χ_{t+1}), expressed as surprise, has an indirect influence on subsequent brain states, meaning that consciousness has a monistic, deterministic, and functional role in brain states.

Consciousness, in short, is a product of the functions of the brain. When focused on the solitary brain, consciousness is left with no more function than the output display on a computer screen, playing no role in determining the succeeding state of the computer that produced this output. Of course the display appears on a computer screen because such computers are not designed to operate in isolation but rather are designed to make it possible for humans to interact with them. When a human operator is sitting at the keyboard, the computer display plays a crucial role in determining what are the subsequent states of the computing device as it influences what the human operator instructs the computer to do next. Eliminate the computer display, and the subsequent states of the computer are quite different than they would be if the human operator were able to see the computer display. We have posited an analogous argument for the functional yet deterministic and monistic role of consciousness – an argument that follows from the observation that we are a fundamentally social species, which is to say that our brains have evolved to interact with other brains.

6. Conclusion

The premise of social neuroscience is that the study of the human brain is incomplete when treated as a solitary organ. When the brain is viewed within a social context — that is, within a context of interacting brains — the material mechanism by which consciousness plays a role in subsequent brain states is revealed. From the perspective of social neuroscience, language evolved to communicate and coordinate with conspecifics rather than simply to talk to oneself. When communicating with others, our brain spontaneously attempts to predict



the present and future behavior of others through processes following under terms such as social cognition, mentalizing, and theory of mind. Although perhaps better than chance, we are far from perfect mind-readers. This imperfection means that when we communicate our conscious states to others, our subsequent brain states are not entirely predicted by the brain states that gave rise to our conscious expression. That is, our subsequent brain states are in part determined by our prior brain states and in part by the brain states that result from interactions with other brains – other brains whose influence on our subsequent brain states are themselves determined in part by the communication of aspects of our consciousness. One might counter that a person's communication is itself the result of prior and ongoing brain states so that it is the brain states, and not aspects of consciousness, that are being communicated. The case of the display screen on a computer helps clarify why this output is instrumental even if its influence is mediated through interactions with other humans. Thus, because our brain underlies communication with other brains – that is, because it is social, the conscious beliefs and intentions we communicate to others has an impact on the brains of others that were not entirely predictable by their prior brain states, and their responses to us influence our subsequent brain states in ways not entirely predictable by our prior brain states. Consciousness in this social context may therefore have the potential to serve a functional role.

7. Acknowledgements

This research was supported by National Institute of Aging Grant No. RO1 AG034052-01.

References

- 1 Petty, R. E., and J.T. Cacioppo (1981), *Attitudes and persuasion-classic and contemporary approaches* (W. C. Brown Co. Publishers: Dubuque, IA).
- 2 Öhman, A., A. Flykt, and D. Lundqvist (2000), "Unconscious emotions: Evolutionary perspectives, psychophysiological data, and neuropsychological mechanisms", in: *Cognitive neuroscience of emotion*, Eds. R. Lane & L. Nadel (O.U.P.: New York), pp. 296-327.
- 3 Dennett, D. C. (1991), *Consciousness explained* (Little, Brown & Co. USA).
- 4 Grau, J., E. Crown, A. Ferguson, S. Washburn, M. Hook, and R. Miranda (2006), "Instrumental learning within the spinal cord: underlying mechanisms and implications for recovery after injury", *Behavioral and cognitive neuroscience reviews*, 5(4), 191-239.
- 5 Berntson, G. G., J.T. Cacioppo, and K.S. Quigley (1993), "Cardiac psychophysiology and autonomic space in humans: empirical perspectives and conceptual implications", *Psychological bulletin*, 114(2), 296-322.
- 6 Sherrington, S. C. S. (1905), *On reciprocal innervation of antagonistic muscles – Seventh-Eighth note* (Harrison: London).



- 7 Berntson, G.G., and J.T. Cacioppo (2008), "The neuroevolution of motivation", in: Handbook of Motivation Science, eds. J. Shah and W. Gardner (New York: Guilford Press), pp. 188-200.
- 8 Berntson, G. G. and J.T. Cacioppo (2008), "The functional neuroarchitecture of evaluative processes", in: Handbook of Approach and Avoidance Motivation, ed. A.J. Elliot (New York: Psychology Press), pp. 307-321.
- 9 Soon, C. S., M. Brass, H.-J. Heinze and J.-D. Haynes (2008), "Unconscious determinants of free decisions in the human brain", *Nature Neuroscience*, 11, pp. 543–545.
- 10 Shiffrin, R. M., and W. Schneider (1984), "Automatic and controlled processing revisited", *Psychological review*, 91(2), 269-276.
- 11 Edgley, S. A., J.A. Eyre, R.N. Lemon, and S. Miller (1997), "Comparison of activation of corticospinal neurons and spinal motor neurons by magnetic and electrical transcranial stimulation in the lumbosacral cord of the anaesthetized monkey", *Brain: a journal of neurology*, 120(5), 839-853.
- 12 Aron, A.R. (2008), "Progress in executive-function research: From tasks to functions to regions to networks", *Current Directions in Psychological Science*, 17, 124–129.
- 13 Ajzen, I. (1991), "The theory of planned behavior", *Organizational behavior and human decision processes*, 50(2), 179-211.
- 14 Petty, R. E. and J. T. Cacioppo (1986), *Communication and persuasion: central and peripheral routes to attitude change* (Springer-Verlag: New York).
- 15 Nosek, B. A., A.G. Greenwald, and M.R. Banaji (2007), *The implicit association test at age 7: A methodological and conceptual review* (Psychology Press: New York), pp. 265-292.
- 16 Phelps, E. A., K.J. O'Connor, W.A. Cunningham, et al. (2000), "Performance on indirect measures of race evaluation predicts amygdala activation", *Journal of cognitive neuroscience*, 12(5), 729-38.
- 17 Kircher, T., and A.S. David (2003), *The self in neuroscience and psychiatry*. (C.U.P.: Cambridge UK).
- 18 Pinker, S. (2003), "Language as an adaptation to the cognitive niche", in: M. Christiansen & S. Kirby (eds.), *Language evolution: States of the Art* (New York: Oxford Univ. Press).
- 19 Gazzaniga, M. S. (1995), "Principles of human brain organization derived from split-brain studies", *Neuron*, 14, pp. 217–228.
- 20 Cacioppo, J. T., and G.G. Berntson (1992), "Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis", *American Psychologist*, 47, 1019-1028.
- 21 Dunbar, R. (2003), "Evolution of the Social Brain", *Science*, 302(5648), pp. 1160-1161.
- 22 Dunbar, R. I. M. (2009), "The social brain hypothesis and its implications for social evolution", *Annals of human biology*, 36(5), 562-572.
- 23 Dunbar, R. I., and S. Shultz (2007), "Evolution in the social brain", *Science*, 317(5843), 1344-7.
- 24 Chicago Social Brain Network (2011), *Invisible forces and powerful beliefs: Gravity, gods, and minds* (Upper Saddle River, NJ: FT Press).



- 25 Cacioppo, J. T., and B. Patrick (2008), *Loneliness: Human nature and the need for social connection* (New York: W. W. Norton & Company).
- 26 Cacioppo, J. T., and J. Decety (2011), "Social neuroscience: Challenges and opportunities in the study of complex behavior", *Annals of the New York Academy of Sciences*, 1224, 162-173.
- 27 Gregory, R. L. (1987), *Oxford Companion to the Mind* (O.U.P.: Oxford, UK).
- 28 Haggard, P., and M. Tsakiris (2009), "The experience of agency: Feelings, judgments, and responsibility", *Current Directions in Psychological Science*, 18, 242-246.
- 29 Herrmann, E., et al. (2007), "Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis", *Science*, 317, 1360-1366.
- 30 Nosek, B. A., A.G. Greenwald, and M.R. Banaji (2007), *The implicit association test at age 7: A methodological and conceptual review* (Psychology Press: New York), pp. 265-292.
- 31 Porter, R. (1987), "Functional studies of motor cortex", *Ciba Foundation Symposium*, 132, 83-97.
- 32 Povinelli, D. J., H.K. Perilloux, J.E. Reaux, and D.T. Bierschwale (1998), "Young and juvenile chimpanzees (*Pan troglodytes*) reactions to intentional versus accidental and inadvertent actions", *Behavioural Processes*, 42, 205-218.
- 33 Rescorla, R. A. & A. R. Wagner (1972), "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement", in: A.H. Black & W.F. Prokasy (eds.), *Classical Conditioning II: Current Research and Theory* (New York: Appleton-Century-Crofts).
- 34 Rinn, W. E. (1984), "The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions", *Psychological bulletin*, 95(1), 52-77.
- 35 Shultz, S., and R. I. M. Dunbar (2006), "Both social and ecological factors predict ungulate brain size", *Proceedings of Biological sciences from The Royal Society*, 273(1583), 207-215.
- 36 Silk, J.B., et al. (2003), "Social bonds of female baboons enhance infant survival", *Science*, 302, 1231-1234.