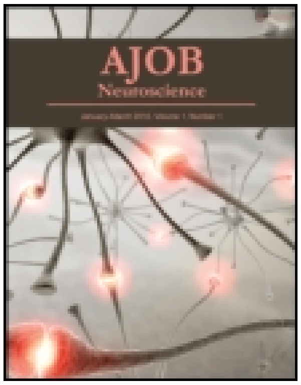

AJOB Neuroscience
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uabn20

How the Neuroscience of Decision Making Informs Our Conception of Autonomy

Gidon Felsen [a] & Peter B. Reiner [b]
[a] University of Colorado School of Medicine
[b] University of British Columbia

Available online: 05 Jul 2011

To cite this article: Gidon Felsen & Peter B. Reiner (2011): How the Neuroscience of Decision Making Informs Our Conception of Autonomy, AJOB Neuroscience, 2:3, 3-14

To link to this article: http://dx.doi.org/10.1080/21507740.2011.580489

AJOB Neuroscience, 2(3): 3–14, 2011
Copyright © Taylor & Francis Group, LLC
ISSN: 2150-7740 print / 2150-7759 online
DOI: 10.1080/21507740.2011.580489

Target Article

How the Neuroscience of Decision Making Informs Our Conception of Autonomy

Gidon Felsen, University of Colorado School of Medicine
Peter B. Reiner, University of British Columbia

Autonomy, the ability to make decisions for ourselves about ourselves, is among the most prized of human liberties. In this review we reconsider the key conditions necessary for autonomous decision making, long debated by moral philosophers and ethicists, in light of current neuroscientific evidence. The most widely accepted criteria for autonomy are that decisions are made by a rationally deliberative and reflective agent and that these decisions are free of undue external influences. The corpus of neuroscientific data suggest that human brains are capable of the hierarchical control required for reflective thought, but that decisions conventionally perceived as autonomous may not be rational with respect to the deliberative process itself, and are rarely free from covert external influences. These findings cast doubt upon the capacity for autonomy as traditionally defined, and suggest that we reconsider valorizing the right to autonomy in order to align our moral values with neuroscientific naturalism.

Keywords: decision making, neuroethics, neuroscience

The nature of autonomy has long been a matter of philosophical debate. What makes a decision autonomous? Who (or what) is capable of deciding and acting autonomously? To what extent should individual autonomy be valued? These questions are significant because autonomy is considered to be a precondition for moral agency: Our conception of autonomy has important implications for personal responsibility and the relationship between the individual and society, including whether it is permissible (or even advisable) for the state to influence individuals' choices for the common good. The utilitarian and the Kantian perspectives on these questions each depend upon particular assumptions about how we make decisions. However, the philosophical conception of autonomy has been formulated primarily introspectively, incorporating ideas from folk psychology with little input from empirical studies of decision making.

While decisions have traditionally been studied at the behavioral level within the disciplines of psychology and economics, advances in the neurosciences have begun to shed light upon the neural mechanisms of decision making (Gold and Shadlen 2007). Given the close relationship between autonomy and how we make decisions, a modern conception of autonomy would benefit from incorporating the relevant findings from neuroscience. In this paper, we critically examine the conception of autonomy from the perspective of neuroscience, with the goal of grounding the debate with empirical neurobiological evidence.

A CONSENSUS CONCEPTION OF AUTONOMY

As is the case with many concepts, autonomy is notoriously difficult to define with precision (Dworkin 1988; Christman 1989; Rosch 1999; Mackenzie and Stoljar 2000). Autonomy literally means "self-rule," and can be used in several context-dependent manners. Most broadly, autonomy can be considered (1) the *right* to be free to self-govern, and (2) both the *state* of being capable of, and actually exercising, self-government (Christman 1989). These concepts are interrelated, since it would be worthless to have the right to autonomy without actually being able to exercise it, and it may be impractical to exercise autonomy without having the right to do so. The philosopher of mind and the criminologist are concerned with the state of autonomy, since it determines the control the individual has over her or his decisions and thus her or his degree of responsibility for them. The political libertarian invokes the right to be free from interference with one's own decisions. In the context of clinical bioethics, the emphasis on informed consent demonstrates the attempt to preserve the right to, and implicitly assumes the state of, autonomy. For our purpose of examining the neuroscientific underpinnings of autonomy, we are primarily, but not exclusively, concerned with the necessary conditions for its state. Specifically, what features must a decision have in order for it to be considered autonomous? With the caveat that the rich philosophical debate is ongoing, we believe that the following three conditions can be distilled from it, constituting what we will

Address correspondence to Dr. Gidon Felsen, University of Colorado School of Medicine, Physiology & Biophysics, 12800 E. 19th Ave., Mail Stop 8307, Aurora, CO 80045, USA. E-mail: Gidon.Felsen@ucdenver.edu

call the "standard model" of autonomy. For a decision to be autonomous, it must be:

1. Consistent with the individual's "higher-order" beliefs and desires (Frankfurt 1971; Dworkin 1988). It does not suffice to decide to satisfy a "lower-order" physiological need, without reflection at, and "authentication" from, higher levels.
2. Rational: It must be dispassionate, based on explicit information, and be allowed sufficient time for the option with the largest subjective utility to be selected (Christman 1991; Sugden 1991).
3. Not unduly influenced by external factors beyond the individual's control. While complete independence may not be required, covert influences on decisions would pose a threat to autonomy (Dworkin 1976).

At first glance, these criteria would appear to deem some decisions autonomous and others not. Choosing a career path, under ideal conditions, may meet all of the criteria and therefore be an autonomous decision. Reacting to a sudden threat, without time to reflect on the response options, would not. As we hope to convey in this paper, most decisions are much more difficult to classify as either fully autonomous or non-autonomous, both because the concept itself is not clearly defined and because the evidence from neuroscience is unlikely to provide a clear line of demarcation (Muller and Walter 2010).

The preceding conditions are not mutually exclusive. For example, a decision consistent with the individual's higher-order beliefs is often free from undue external influence as well. And certainly, other notions of what the term means have been proposed, and these have merit in and of themselves (for comprehensive reviews, see Christman 1989; Christman and Anderson 2005; Taylor 2005). To take but one example from the many thoughtful offerings in the field, Christman (1991) proposes that autonomy with respect to some desire requires the individual to have not resisted the development of that desire. Nonetheless, the standard model represents the consensus of the core of autonomous decisions, and has been endorsed in some form by the preponderance of investigators, including the intellectual forebears of this view of autonomy. Kant, for instance, suggested that autonomous decisions are reasoned, based on principles rather than (first-order) desires, and "independent of alien causes" (Hill 1989; Kant et al. 2002).

While understanding the neurobiology of how decisions are made can shed light on autonomy, we make no claim that it will provide all of the answers we seek. In particular, normative questions about autonomy are best addressed within the discipline of moral philosophy. However, any prescription for how things *ought* to be rests on the assumption that things *can* be that way (Kant et al. 1999). Thus, what neuroscience can provide is a naturalistic framework within which to ground, inform, and constrain the philosophical debate. For example, the value of a philosophical viewpoint that depends upon a particular capacity would be limited by evidence from neuroscience inconsistent with that capacity. With this goal in mind, we now discuss the neuroscientific evidence for and against autonomous decision making.

THE NEUROSCIENTIFIC BASES FOR THE STANDARD MODEL OF AUTONOMY

Hierarchy of Desires

Dworkin posits that "a crucial feature of persons [is] their ability to reflect upon and adopt attitudes toward their first-order desires, wishes, [and] intentions" (Dworkin 1988). Similarly, Frankfurt states that "It seems to be peculiarly characteristic of humans. .. that they are able to form what I shall call 'second-order desires'" (Frankfurt 1971). "First-order" desires often—but do not always—correspond to physiological needs, such as hunger, and to reflexive emotions, such as resentment at being mistreated. The individual's "second-order" desires, or volitions, comprise his attitudes about his first-order desires, such as whether or not he wants to have them. Autonomy, then, is "conceived of as a second-order capacity of persons to reflect critically upon their first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values" (Dworkin 1988).

Frankfurt takes a similar view: An autonomous individual is "free to want what he wants to want" (Frankfurt 1971), as opposed to being enslaved to his or her first-order desires. Positing multiple levels of desires represents an attempt to differentiate decisions consistent with the individual's fundamental goals and beliefs, which are considered autonomous, from other decisions, which are not. For example, an addict who decides to satisfy a first-order craving for a drug, despite a second-order desire to not have that craving, would not be considered to have decided autonomously. Critically, both Dworkin and Frankfurt consider the capacity for autonomy to be uniquely human.

Reflections upon second-order desires could be considered third-order desires, reflections upon those would be fourth-order, and so on. While classifying a desire as being of one particular order may be overly simplistic, the general structure forms the basis of a rich hierarchy of desires, which is analogous in many ways to the organization of neural structure and function. The gross anatomy of the brain is largely (but not exclusively) hierarchically organized. In general, the brainstem, the evolutionarily oldest part of the brain, is responsible for unconscious, automatic behaviors, such as the regulation of breathing and body temperature. The cerebral cortex mediates more complex cognitive tasks, which often requires conscious awareness. Information can be processed in a "bottom-up" fashion, flowing from the brainstem through midbrain nuclei and to the cortex, and "top-down," in which the cortex modulates brainstem activity. This account is of course oversimplified; the brain consists of many interconnected subregions responsible for processing and representing specific information. To a first approximation, however, lower- and higher-order desires can be thought to map onto representations in the brainstem and cortex, respectively, and it is generally accepted that the cortex is at the top of the hierarchy for autonomous decision making.[1]

Consistent with this notion, there exists a strong correlation between cortical function and the capacity for autonomous decisions. Animals with a comparatively rudimentary cortex (e.g., birds and reptiles), and humans with poorly developed cortical function (e.g., infants and the mentally disabled) or cortical damage (e.g., due to injury or stroke) are generally not considered to be fully capable of autonomous decision making. Moreover, even healthy human adults may transiently lack autonomy under heightened emotional conditions during which decisions may not be under cortical control (see further discussion).

The hierarchical theory of autonomy requires not only that there exist multiple levels of desires, but also that autonomous decisions actually reflect the higher-order desire. How could this organization be realized in the brain? One possibility is suggested by executive control theory (ECT), which posits that the prefrontal cortex (PFC), the cortical region most fully developed in humans, exerts "top-down" influence over other brain regions (Miller and Cohen 2001). The idea is that persistent activity in local circuits of PFC neurons, representing behavioral goals, is thought to bias the computations performed in other brain regions such that their ultimate output is consistent with those goals. For example, consider being in line at a cafeteria and deciding whether to have a salad or a brownie. Activity in distinct circuits would represent the sequence of actions required to select one or the other. These two circuits "compete" with each other, such that one of them "wins" and determines the behavioral output (e.g., eating the salad). The PFC can bias this competition such that its outcome is consistent with the goals it is currently representing. If that goal is to avoid gaining weight, the PFC would make it more likely that the "salad circuit" wins.

This example highlights several key features of the neural correlates of a hierarchy of desires. First, the higher-order desires correspond to the behavioral goals represented in the PFC—in this case, to avoid weight gain—while the lower-order desires correspond to wanting the brownie and wanting the salad. Note that the lower-order desires correspond to the physiological need for food, while the higher-order desire represents an attitude about the lower-order desires (e.g., "I do not *want* to want a brownie"): a meta-desire, if you will. Second, decisions to satisfy physiological needs are often unconscious, despite the complex sequence of motor action required, as anyone who has "absentmindedly" eaten an entire batch of brownies can attest. Higher-order desires, on the other hand, are more likely to be (but are not necessarily) represented in conscious awareness. This notion fits well with the simplified idea of conscious cortical, and unconscious subcortical, processing. Third, while higher-order desires may make it more or less likely that a particular lower-order desire will be satisfied, they do not always guarantee the outcome. The decision is thus determined by the relative strengths of the lower-order and higher-order desires: Despite the identical desire to avoid gaining weight, you might choose a particularly tasty brownie (e.g., with chocolate frosting) over the salad. The ECT accounts for this situation by positing that the brownie circuit has an advantage that even the bias from the PFC cannot overcome. Fourth, lower-order desires are generally focused on immediate rewards, while higher-order desires are more sensitive to long-term outcomes. Finally, note that higher-order desires are not necessarily in the individual's best interest. For example, societal influences may lead one to highly value thinness, resulting in a desire to lose as much weight as possible, even at the expense of overall health.

An important difference exists between the representation of higher-order desires and behavioral goals in the brain. High-level desires may be very long-lasting; for example, people may carry the desire to be healthy with them throughout their lifetime. However, activity representing behavioral goals persists for minutes or hours at most (Fuster 1995; Goldman-Rakic 1995). The neurobiological canon would hold that high-level desires are stored as long-term memories, for example, in the synaptic weights of hippocampal and cortical networks (Hebb 1949; Bliss and Lomo 1973; McNaughton and Morris 1987). When required, these representations are transiently transformed into persistent PFC activity and are used to guide decisions.

Early evidence for the idea that the PFC performs executive control functions arose from human lesion studies. In particular, it was observed that PFC damage specifically disrupted higher-order behavioral control without affecting the ability to perform simple tasks (Milner 1963). These results suggested that the PFC was specifically involved in coordinating behavior. Evidence for an executive function for activity in local PFC circuits came from electrophysiological recordings in primates demonstrating persistent activity during behavioral tasks (Fuster and Alexander 1973), including activity that explicitly represents contextual information required for achieving behavioral goals (Romo, Brody et al. 1999; Wallis, Anderson et al. 2001). More recently, imaging studies in humans have demonstrated distinct roles played by subregions of the PFC in top-down control, such as anterior regions of the PFC controlling those more posterior (Badre 2008; Badre and D'Esposito 2009). Together, these data support the main features of the ECT: the functional relevance of persistent activity in the PFC for biasing processing in other circuits in order to control behavior.

The hierarchical organization proposed by the ECT endows the nervous system with a powerful mechanism for behavioral control. Instead of producing stereotyped responses to the same input, the hierarchy provides the nervous system with the capacity for behavioral flexibility. Such flexibility is adaptive for the organism's survival insofar as it permits the exploration of new behavioral strategies and expands the range of environments in which the organism

[1] While we primarily present data supporting the idea that the prefrontal cortex sits atop the hierarchy, alternative viewpoints exist. The specific brain regions constituting each level of the hierarchy are not critical for the purposes of our discussion.

can thrive. Presumably, then, flexibility in decisions and behavior has been selected for by evolution. The potential for autonomy, which is also dependent on hierarchical organization, may have evolved in parallel with the capacity for flexibility in decisions and behavior. This idea is consistent with the notion that organisms with less well-developed cortical function tend to exhibit a lower degree of autonomy.

However, there is a fundamental conceptual difficulty with the hierarchical model, with respect to both brain function and autonomy: What happens at the top of the hierarchy? In neuroscience, this problem is exemplified by invoking the existence of a homunculus: a "little man" found at the apex that is the ultimate controller. Of course, this is not a viable solution, since it simply passes on the question to what is controlling the homunculus. A similar "regress problem" also occurs with respect to the hierarchy of desires (Watson 1975). With respect to the salad versus brownie decision discussed earlier, perhaps the second-order desire to avoid gaining weight is influenced by a third-order desire to maintain good health, which answers to a fourth-order desire to live as long as possible, and so on. If autonomy requires authentication from a higher-order desire, how do we avoid the trap of an infinite number of orders of desire?

Although neuroscience has not provided resolution to the dilemma posed by the top of the hierarchy, several theories have been proposed. One suggestion extends the idea that the dopaminergic system is responsible for reinforcing rewarded actions by applying the same role to the patterns of persistent PFC activity themselves: those patterns—patterns that we suggest correspond to higher-order desires—that lead to positive outcomes are reinforced, and are thus more likely to occur in the future (Braver and Cohen 2000; Hazy, Frank et al. 2007). One might also posit a variant of the multiple drafts model that has been proposed for consciousness (Dennett 1991): that decisions occur not at any one locus or even instant, but rather arise through parallel, multiplexed information processing. Our introspective *perception* of the decision as occurring at a given time leads us to conclude that there is a single brain region responsible for the decision, but in principle the computations required could just as easily be accomplished via a distributed network of synaptic events. Regardless of the specific mechanism, the PFC is a leading candidate to have evolved the ability to reflect upon, evaluate, and select among the lower-order desires represented in the activity of neural circuits.

Several solutions have also been proposed for the regress problem. One appealing suggestion posits that an original autonomous and authentic structure gradually emerges during development, incorporating features into its self-representation as new situations are experienced and decisions become associated with their outcomes (Hofstadter 2007). This structure can then confer authenticity on other needs and desires, and need not (in fact, cannot) be authenticated itself, since the infant brain is not capable of autonomous decisions (Noggle 2005). An obvious question then is when, in the development from infancy to adulthood, the hierarchical control necessary for autonomy first emerges. We concur with the view that the capacity for multiple orders of reflection, abstraction, and representation may distinguish adult human cognition from that of infants and nonhuman animals (Hofstadter 2007). Although identifying such a time point is desirable from an ethical and legal perspective, it is not clear that even a perfect understanding of the underlying neuroscience could provide one. Instead, the neuroscientific evidence supports the notion of experience-dependent degrees of hierarchical control, spanning the range from infant (none) to healthy adult (full). This idea is consistent with current thinking on the rights and responsibilities of children with respect to their own medical decisions: that these should depend on experience (e.g., previous treatment history) rather than age (Alderson 2007), and that children may "assent" (agree) to decisions made on their behalf, rather than offer fully informed consent (Committee on Bioethics 1995). Similarly, the perspective from neuroscience supports the idea that while humans have the greatest capacity for autonomy, other species may exhibit varying degrees of autonomy, contrary to the view espoused by Dworkin and Frankfurt, perhaps depending on the particular context for the decision.

The idea that autonomous decisions require authentication from higher-order desires is paralleled by the hierarchical nature of neural control postulated by the ECT. While numerous details remain to be determined, and we do not claim that specific levels of desire correspond precisely to particular brain regions, the ECT provides a framework for how decisions and actions may be influenced by higher-order desires. We therefore find good neurobiological support for the viability of decisions being controlled by a hierarchy of desires, an important component of the standard model for autonomous decision making.

Rationality

The second requirement of the standard model for autonomous decisions is that they are rational. While the notion of rationality has long been debated, the traditional view of rational decision making focuses on the deliberative process itself. Specifically, the process requires access to all of the relevant information and sufficient time and neural resources to select the option associated with the best predicted outcome: the advantages and disadvantages of each option are weighed, and the most valuable one is selected (Becker 1976; Sugden 1991; cf. "rationality$_2$" of Evans 2003). There is much evidence demonstrating that humans and animals are capable of weighing options and choosing the optimal one. Behavioral economics has successfully explained choices in terms of their subjective value (Kable and Glimcher 2009), and game theory has demonstrated that animals and humans are capable of adopting a seemingly optimal strategy in order to maximize their reward (Von Neumann et al. 1947; Glimcher and Rustichini 2004). Recently, many studies have analyzed the neural substrates of economic decision making, demonstrating where and how value is represented in the brain and how these representations are used to guide decisions (Montague and Berns 2002; Glimcher and Rustichini 2004; Sugrue et al. 2005; Gold and Shadlen 2007). Further, several cortical regions are engaged by reasoning and deliberation, consistent with their role in executive control (Jung and Haier 2007).

While these studies identify a *capacity* for rational decision making and its neurobiological mechanisms, they do not necessarily address how particular real-life decisions actually *are* made. It had long been intuited that our decisions are influenced by factors of which we are not aware, but it was not until the emergence of prospect theory that evidence accrued demonstrating that behavior predictably deviates from subjective utility theory (Tversky and Kahneman 1981; Ariely 2008). Indeed, many decisions important for survival—how to react to the immediate threat of a predator, for example—must be made without the cumbersome computations required to explicitly consider all of the information and options. But implicit considerations such as these, relying upon covert, partial information, violate a requirement of rationality. In short, much evidence suggests that there are many instances in which we do not make rational decisions.

While the specific neurobiological mechanisms by which decisions are covertly influenced are not yet fully understood, one theory has attempted to explain how emotions, even those outside of conscious awareness, contribute to decision making. This theory, known as the somatic marker hypothesis (SMH) (Damasio 1994; Damasio 1996), suggests that specific networks in the brain have evolved so as to utilize the (nonrational) information generated by emotional responses in order to improve decision making. The idea is that the nervous system "tags" (marks) the outcome of particular choices with particular body (somatic) states that can then be used to guide future decisions. These markers develop through learning the correspondence between the value of a given course of action (i.e., its predicted future outcome) and the emotions associated with that course of action. A negative marker decreases the likelihood that the option associated with it will be selected, while a positive marker increases the likelihood. For example, the thought of eating an entire batch of brownies, while rewarding in the short term, may elicit wide-ranging negative feelings overall (e.g., guilt), making you less likely to do it. Resisting the brownies may elicit emotions associated with a positive outcome in the long term (e.g., pride at maintaining a healthy lifestyle). Such associations between emotions and outcomes must be learned through experience: The first time you had the opportunity to eat all of the brownies, you probably did (notably, your pet dog would probably never "think twice" about doing so). Somatic markers increase the efficiency of the decision-making process by rapidly eliminating options that are obviously undesirable, and by biasing the decision toward or away from particular options depending on their associated emotions.

Since the SMH requires the integration of somatic signals, emotions, outcome representations, memories, and motor output, several brain regions are thought to be involved, and an explanatory framework for how and where the information is processed has emerged. For example, the limbic system, and in particular the amygdala, is necessary for processing instinctive emotions (e.g., fear). Without normal output from this system, it would be impossible to associate emotions with particular situations. The somatosensory cortex is responsible for representing bodily states. The PFC, due to its functional connectivity with all of these systems and others responsible for the already-described functions, and its dense recurrent connectivity, provides an ideal substrate for forming the associations between the actions, values, and somatic markers that are critical to the SMH.

Early support for the SMH, and for the importance of the PFC in producing the appropriate somatic states for particular situations, emerged from studies of individuals with frontal lobe damage. While these patients produced normal autonomic activity in response to simple stimuli, they failed to produce learned responses to more complex stimuli (Damasio et al. 1990). These findings were consistent with early reports of the effect of frontal lobe damage, after which a specific deficit in producing proper emotions was observed despite no apparent change in general intelligence (Harlow 1868). In order to study how this deficit in forming somatic markers affected decision making, the Iowa Gambling Task was developed (Bechara et al. 1997). In each trial of this task, human subjects select a card from one of four decks. Cards from two "good" decks offer either small monetary gains or small losses, while those from two "bad" decks offer either large gains or large losses. Over the course of the session, the subject is likely to gain more by selecting from the "good" decks. Interestingly, while normal subjects quickly learned the optimal strategy, they began to exhibit elevated skin conductance responses (indicating heightened emotional processing) during picks from bad decks sooner than they reported being consciously aware that the deck was bad. Furthermore, patients with bilateral damage to the ventromedial PFC neither exhibited elevated skin conductance responses nor learned the optimal strategy. These results suggest that emotional processing, even when covert, is beneficial to optimal decision making, and that the ventromedial PFC is necessary for this processing. More recent studies have found specific roles for several other PFC regions as well (Fellows and Farah 2005; Dunn et al. 2006). Indeed, the SMH can perhaps best be considered as a special case of the hierarchical ECT: Both propose systems of top-down control, originating with representations in the prefrontal cortex which serve to bias decisions.

The emotional influence on decision making postulated by the SMH can be thought of as a more elaborated form of the primitive "fight-or-flight" response exhibited throughout the animal kingdom. This response enables the nervous system to make rapid survival-promoting decisions when confronted with dangerous situations: With very little deliberation required, the stimuli directly trigger the motor output necessary for either escaping or confronting the danger. The SMH extends this idea in two ways. First, it suggests that not only do instinctive emotions influence decisions (as they do during fight-or-flight), but secondary emotions, which are learned through trial and error, do so as well. Second, it posits that these emotions change the likelihood of particular decisions being undertaken, rather than being "hardwired" to produce a particular stereotyped response.

Just as the fight-or-flight response is evolutionarily advantageous, so too are the associations between secondary emotions and decisions.

It is important to note that the SMH does not suggest that emotional and rational decision making are incompatible. In Damasio's words, "The action of biological drives, body states, and emotions may be an indispensable foundation for rationality" (Damasio 1994, 200). In particular, emotional decision making may act in a "cognitively economical" manner by narrowing the set of options on which the computationally expensive value-based processing just described must be performed in order to select the "best" option. Although this initial winnowing stage would not traditionally be considered rational because it eschews deliberation, the outcome (i.e., the ultimate decision) may match the result of a truly deliberative process. Indeed, somatic markers are only advantageous to the extent that they provide an efficient heuristic for making adaptive decisions (Marewski et al. 2010). If we allow that a rational decision could be produced by a previously learned "shortcut" yielding an identical outcome to a deliberative process, a premise that might be termed *neurobiological consequentialism*, emotions would not pose a threat to rationality, but rather may provide an efficient mechanism for it (Frank 1988; De Sousa 1990; Evans et al. 1993; Chase et al. 1998). For example, in the experiments just described, the "normal" subject is able to utilize unconscious somatic cues to rapidly identify the bad decks of cards. He or she could have arrived at the same decision after calculating the probabilities of gains and losses associated with each deck, but that approach would have taken far longer. Even complex moral decision making may rely more on emotions than on logical reasoning (Haidt 2001).

Of course, emotional decision making may not always be consistent with rationality; in some cases reliance on somatic markers may lead to a very different decision than would a traditionally rational, deliberative process. In post-traumatic stress disorder, for example, negative somatic markers become associated with the set of stimuli present at the time of the stressful event, resulting in counterproductive behavioral responses to those stimuli in the future. Further, in situations in which complete information is available, as when a gambler counts cards, purely logical decisions may be hindered by the presence of somatic markers associated with loss aversion (Shiv et al. 2005; Kuo et al. 2009). We thus suggest that, within this alternative framework in which rationality depends upon the outcome of the decision, the neuroscientific evidence supports the notion that individuals are capable of making, and often do make, rational decisions.

Freedom From Undue Influences

The final condition for autonomous decisions is that they must be free from undue influences external to the individual's collection of higher-order beliefs and desires. There exist two broad classes of such influences: those that are internally generated but are not part of the higher-order beliefs of the individual (such as biological drives), and those arising from the individual's environment. While the former influences clearly affect our decisions, because they originate from "within," it is reasonable to posit that they do not compromise autonomy to any substantial degree (for the sake of simplicity, we exclude pathological states that compromise the brain, although it is clear that such events pose daunting challenges for autonomy [Gleichgerrcht et al. 2010]). We are more interested in what neuroscience can tell us about the ability of external influences to compromise autonomy, for this is really the concern that suffuses the literature. Of course, few (if any) decisions are entirely independent of the individual's family, culture, and other social groups. The relevant question, then, is what sort of external influences do represent a threat to autonomy? As discussed earlier, for a decision to be "one's own," it must be *consistent with* one's higher-order desires (Dworkin 1976). It follows that external factors that can be consciously incorporated into one's decision do not pose a problem for autonomy. However, covert influences, which act unconsciously, potentially subvert the individual's higher-order desires and thus would constitute such a threat.

There are abundant data demonstrating that decisions are influenced covertly. For example, it has long been known that subjects are more likely to (seemingly spontaneously) produce a word, image, or even a concept to which they had been previously exposed, even if they do not explicitly remember the exposure (Tulving and Schacter 1990). It is easy to see how "priming" a subject in this manner can covertly bias future decisions. In one remarkable demonstration, subjects surreptitiously exposed to words associated with a stereotype of the elderly, such as "wrinkle," "Florida," and "conservative," tended to walk more slowly after leaving (what they thought was) the experiment than control subjects exposed to neutral words (Bargh et al. 1996). Although it is difficult to rule out other explanations, the authors accounted for their results by proposing that the experimental subjects had been covertly primed with the concept of "elderly," which unconsciously influenced their subsequent behavior.

The neurobiological explanation of this effect is that the initial exposure to the primed object alters its neural representation in such a way that it is more easily triggered when it is subsequently considered during the decision. This explanation is reminiscent of the biasing of neural activity produced by the PFC that is hypothesized by the ECT, but occurs unconsciously. One potential mechanism, which has been demonstrated at both the single-cell level in animals and the population level in humans, is response suppression (Desimone 1996; Wiggs and Martin 1998), whereby repeated presentation of stimuli results in a decreased neural response. The idea is that the stimulus representation is sparser (i.e., represented by the activity of fewer neurons), and therefore "sharpened," by previous exposure, making it more likely to be accessed by the subsequent decision. It has also been proposed that the bias can occur via modification of top-down processes: The same mechanisms that the PFC normally employs to bias choices are exploited by the prior exposure in order to bias response selection (Schacter et al. 2007). Regardless of the specific mechanism, what is clear is that the process is not under conscious control.

It has also been shown that the manner in which options are presented can, outside of conscious awareness, influence subsequent decisions. Classic examples of such a "framing" effect demonstrate that individuals tend to be risk-averse when the outcome is presented in terms of gains, but risk-seeking in order to avoid the identical outcome if it is presented in terms of losses (Tversky and Kahneman 1981). For example, subjects' selections among medical treatment options depend upon whether identical outcomes are framed as survival rates (e.g., 90% of patients live) or mortality rates (e.g., 10% of patients die) (McNeil et al. 1982; Armstrong et al. 2002). This logical inconsistency is *practically* important because it can be exploited in order to influence individuals' choices. Indeed, it has been shown that physicians tend to frame the presentation of treatment options in order to elicit the desired decisions from patients (Sullivan et al. 1996; McNeely et al. 1997) and that this strategy is effective (Gurmankin et al. 2002; Covey 2007).

While asymmetric risk behavior under gains and losses has been studied extensively at the behavioral level (Kuhberger 1998), its neural bases have only recently begun to be examined (Trepel et al. 2005). One study imaged neural activity in human subjects performing a gambling task with two types of trials: In "gain" trials, they chose to either keep £20 of £50 that they had previously been given ("sure option") or to gamble to either keep or lose all £50 ("gamble option"); and in "loss" trials, they chose to either lose £30 of the £50 they had been given (sure option) or to take the same gamble as in the gain trials (gamble option) (De Martino et al. 2006). The amygdala was found to be more active during gain trials when subjects chose the sure option and during loss trials when subjects chose the gamble option (which were the most common choices, consistent with previous results; Tversky and Kahneman 1981) than during the other two conditions. These results suggest that, consistent with the SMH, covert emotional processing (as suggested by an increased BOLD signal, a proxy for brain activity, in the amygdala) may play a role in producing divergent behavior under gain and loss frames. Specifically, the fear of losses may have more influence on the decision than the pleasure of gains, which explains why subjects (suboptimally) act to avoid perceived losses by choosing the sure option in gain trials, and the gamble option in loss trials. An alternative explanation was offered by a similar study that found that choices reflecting loss aversion could be predicted based solely on the BOLD activity in the brain regions that process reward gains (Tom et al. 2007). Modeling studies have attempted to reconcile these findings, suggesting that the effects of framing result from interactions among multiple neural systems representing both gains and losses (Litt et al. 2008).

Another example of information framing covertly influencing decisions is known as the "anchoring effect." Here, individuals are willing to pay more for a product if they have first been prompted to think of a high number rather than a low number (Ariely et al. 2003). The value of the product appears to be assigned relative to the arbitrarily determined initial number, consistent with the finding that the nervous system generally represents value in relative, rather than absolute, terms (Seymour and McClure 2008). For example, single neurons in the primate orbitofrontal cortex have been shown to represent the relative value of two options, responding differently to the same cue when it is presented in two contexts (i.e., predicting the higher reward in one context and the lower reward in the other) (Tremblay and Schultz 1999). Encoding relative value makes evolutionary sense: Since values are only useful for deciding between multiple options, it is not adaptive to assign an absolute value to a single arbitrary option. Further, the available range of neuronal firing rates is limited by the metabolic cost of producing action potentials (Laughlin et al. 1998). It would therefore be more efficient to adjust the relationship between value and firing rate to encode the currently relevant range of values, as some sensory neurons have been shown to do (Laughlin 1981). Encoding relative value has been adaptive for decisions important in our evolutionary history, but it can clearly be (ab)used to manipulate decisions in the modern world—for example, to maximize what consumers will pay for a given item.

Product marketing also takes advantage of other tactics for influencing decisions, such as branding. One study replicated the well-known "Coke vs. Pepsi" taste tests while recording subjects' neural activity using functional magnetic resonance imaging (fMRI) (McClure et al. 2004). They found that subjects' opinions and the BOLD signal differed when the subjects knew which beverage they were drinking compared to when they did not. These results demonstrate that expectation, cued by a familiar brand associated with prior experiences, has a significant effect on perception and judgment.

We close our discussion of covert influences by drawing the reader's attention to an idea that arises from philosophy of mind rather than neuroscience. A dozen years ago, Clark and Chalmers famously introduced the extended mind hypothesis (Clark and Chalmers 1998), suggesting that the "mind" extends beyond the corporeal body and includes certain aspects of the environment around us as well. The extended mind hypothesis lends itself quite effortlessly with the conclusion that there exists a panoply of covert influences *out there* over which we have little control. The extended mind hypothesis is hardly without its detractors (Adams and Aizawa 2009; Rupert 2009), and while the neurobiological evidence suggests that the *decision* is made in the brain, there is no clear demarcation with respect to the information that the brain uses to make that decision. In this manner, the extended mind hypothesis can be seen to reinforce the idea that decisions can be influenced covertly by external sources. Since these effects occur outside of conscious awareness, they cannot be neutralized by the rational drivers of our decisions and behavior. It is important to note that these influences are not necessarily malevolent, and are often even encouraged. For example, childhood education, generally considered to be a positive feature of healthy societies, entails the influencing of beliefs and actions. Influencing choices in many other contexts may similarly be employed for the benefit of the individual and society (Thaler and Sunstein 2008). Critically, decisions are also influenced covertly by *internal* sources in ways that may be beneficial to the agent. For example, without being consciously aware of them, signals produced by the circuits that control feeding behavior may influence one's decision such that the individual eats at the appropriate time. Indeed, one reason that covert external factors may be so difficult to avoid is that once they are "internalized," they are indistinguishable from authentic internal sources to other brain areas involved in the decision. Since covert internal influences are often adaptive, external influences can be (mis)interpreted as adaptive as well. Thus, at least some of our decisions do not appear to be as free from undue influence as the standard model of autonomy requires.

DISCUSSION

We have examined the neurobiological bases of the three primary conditions necessary for autonomous decisions: that they are consistent across levels of a hierarchy, rational, and sufficiently free from external influences. We have found that decision making and behavior are well described by hierarchical processes in the brain. However, we also argue that the neuroscientific evidence only supports the idea that our decisions are rational within a fundamentally reconceptualized framework of rationality, and find that many decisions that are traditionally considered autonomous are not necessarily free from covert external influence. This evidence thus calls into question the extent to which the decisions made by healthy human adults are autonomous, at least in the strictest sense.

Does the evidence opposing the traditional conception of autonomy demonstrate some deficiency in brain function? On the contrary, the brain has evolved to optimally perform those functions necessary for survival, including making the decisions leading to the best outcome in a given situation. It may be the case that many decisions can best be made (in part) nonrationally—for example, by using somatic markers in order to avoid explicitly weighing the costs and benefits of every alternative—and thus, non-autonomously. Indeed, purely rational decision making would likely be too time-consuming to be efficient in the real world, and as a result evolution has selected against relying upon it exclusively. As valued as autonomy is in modern society, it was likely not an adaptive feature in the history of our species. There is thus no reason to assume, a priori, that the modern brain would have evolved to support autonomous decision making.

However, we make no claim that any *particular* decision cannot be autonomous. With enough time and effort, an agent is capable of employing deliberative, well-considered, and rational processes in place of rapid, automatic, and emotional processes. For example, an understanding of framing effects can allow them to be overcome: The individual may reframe the same set of options in terms of losses and gains, and use both frames to make a decision. And while it may be difficult to counteract covert priming effects, such influences do not necessarily impinge upon every decision. Further, awareness of such effects may contribute to mitigating them. Thus, the neurobiology does not exclude the possibility that decisions can be made autonomously. Practically, however, what does neuroscience have to say about the autonomy of real-world decisions? To address this question, we briefly apply the principles discussed earlier to decisions made in four contexts in which autonomy has been a topic of debate.

Medical Decision Making

The last several decades have seen a fundamental shift away from the paternalism traditional to the medical establishment toward increased autonomy for patients in deciding on their own medical care (Root-Wolpe 1998). When selecting among multiple treatment options, a patient may exercise autonomy by using her or his higher-order beliefs to guide her or his decisions. For example, although a particular treatment may be painful and thus runs counter to the patient's first-order desire to avoid discomfort, the patient's higher-order desire to remain healthy and able to provide for her or his family may increase the likelihood that the patient will opt to proceed with the treatment. On the other hand, patient autonomy may be compromised in several ways. When presenting treatment options, it may be impossible for the caregiver to avoid framing the choices in such a way that the patient is influenced. Further, the patient's decision is likely to be covertly biased by the fact that the source of the information is a medical expert, and perhaps by unconscious somatic states triggered by potential treatment outcomes. These seemingly unavoidable influences suggest that striving for absolute patient autonomy may be quixotic, and lend support to the practice of "soft paternalism" in which caregivers present options such that they bias patients toward the choices that the caregivers believe to be best, while preserving the patient's ability to select any of the options (Sunstein and Thaler 2003; Swindell et al. 2010). Surveys demonstrate that caregivers regularly adopt a soft paternalism approach (Sullivan et al. 1996; McNeely et al. 1997), perhaps because they implicitly recognize that unqualified patient autonomy is unlikely to be in the patient's best interests, and may be practically unachievable.

Addiction

Addiction is generally considered to compel behavior. For example, the heroin addict is forced to inject the drug in order to relieve his or her craving. While the influence arises from within the individual, it is "external" to his or her higher-order desire to (presumably) be free of the compulsion. Thus, addiction has traditionally been thought to compromise autonomy, if not entirely disable it (Caplan 2008; but see Levy 2006; Buchman and Russell 2009).

Much neuroscientific evidence supports this viewpoint. There are several overlapping mechanisms by which addiction influences behavior, any number of which may operate on a given individual and with respect to a given substance (Hyman et al. 2006; Redish et al. 2008). One experimentally supported theory holds that the midbrain dopaminergic system is "hijacked" by the downstream effects of drugs of abuse (Redish 2004; Volkow and Li 2004). Normally, dopamine is released in response to unexpected rewards or to stimuli that predict future rewards. It is thus thought to serves as a "reward prediction error," signaling the difference between the expected and actual reward obtained (Montague et al. 1996; Schultz et al. 1997). This signal can be used by the nervous system to learn which actions are most likely to lead to reward (Sutton and Barto 1998). However, since dopaminergic signaling is modified by some drugs of abuse, drug consumption is misinterpreted by the reward system as signaling a rewarding event. The actions that led to the reward are thus marked as reward-directed, and become more likely to be selected in the future. The net effect is that each time the drug is taken, the drug-seeking behavior is reinforced, further increasing the frequency of drug taking. This positive feedback cycle ultimately results in a compulsion for drug seeking that can become very difficult to overcome, even if it runs counter to a higher-order desire to avoid the drug. Thus, the neural mechanisms guiding reward-directed decisions can be subverted by addiction to compromise autonomy.

Marketing

Advertisers have long sought to influence consumer behavior with the objective of maximizing product sales. Several such strategies exist, some of which may infringe on consumer autonomy and some of which may not. For example, explicitly providing the information necessary for a shopper to make an informed choice between competing products is no threat to autonomy. However, "stealth marketing," in which information is covertly deployed, poses a real threat (Murphy et al. 2008). In the real world, marketing commonly occupies some middle ground between these extremes, taking advantage of the neural mechanisms that allow decisions to be influenced by factors outside of conscious awareness but without fully hijacking the consumer's decision-making processes. Given the prevalence of marketing that appeals to emotions (e.g., sexual desire) rather than providing neutral information (e.g., the ingredients of the beer being advertised), marketers are clearly aware that this strategy is an effective motivator. With increasing understanding of the neurobiology underlying how covert information affects decisions, infringement on consumer autonomy may become a more serious concern.

Political Philosophy

Our analysis suggests that the notion that one can be confident that any given decision is free of external influence is incompatible with the neurobiological evidence, calling into question the capacity for negative liberty (Berlin et al. 2002). Political philosophies such as libertarianism[2] that exalt negative liberty as the apogee of freedom are thus faced with a dilemma, as the elusive goal that they seek is unlikely to be realized in the human brains acting in the real world (cf. objections to classical liberalism; Christman and Anderson 2005). For negative liberty to be rescued, it may be wise to reconsider the conditions that a decision must meet in order to be regarded as free from external influences. One solution may be to revisit the distinction between external influences that are *authenticated*, at which point they are no longer seen as external but rather as part and parcel of "our own" thinking, and covert external influences that are also incorporated into our thinking but lack authentication. The neuroessentialist perspective, that the *we* that we care about most is *essentially* a collection of conscious and nonconscious brain processes (Reiner 2010), along with the idea that nonconscious mechanisms contribute to executive control (Gillett 2009; Suhler and Churchland 2009; Custers and Aarts 2010), suggests that at least with respect to decision making, there may be no meaningful distinction between those influences that are authenticated and those that are not.

Some may view such a reconceptualization as leaving autonomy so neutered as to be unworthy of value. An alternative approach would be to align our moral philosophy with the neuroscientific evidence. An early champion of this approach was Paul Churchland, who argued for eschewing folk psychology in favor of a philosophical framework that relies primarily on findings from neuroscience (Churchland 1981). Churchland's work is essentially naturalistic, gesturing to the work of Quine (Quine 1969; Flanagan et al. 2007). This thesis recognizes that, due to our evolutionary history as a social species, the human brain is adapted to succeed in social settings and not in isolation. Indeed, human decision making has evolved such that the external influences that comprise our cultural milieu, even those that are covert, usually enable us to act in our own best interests. According to this view, libertarianism as currently envisioned is an unrealistic ideal. It may be possible to learn to make decisions differently, or alter our brain processes in some other way, in order to attain this ideal. But the way our brains actually function, at present, may be an impediment to such a project.

Alternatively, valuing interdependence (Mackenzie and Stoljar 2000) as an ideal to be balanced with individual autonomy in specific contexts may be more in line with the neuroscience of how we make decisions. In promoting this view, some bioethicists have suggested that autonomy need not be considered sacrosanct if respect for it comes at the expense of the public good (Gaylin and Jennings 2003; Turoldo 2009). Although the long-term repercussions of discounting autonomy would need to be examined, if such a policy were effective it could find support, within the framework of scientific naturalism, over policies that valorize the traditional notion of autonomy as the primary determinant of human well-being.

CONCLUSIONS

We have drawn together findings from the neurobiology of decision making in an attempt to understand the extent to which empirical data supports well-established views of autonomy. Our analysis reveals that in many respects,

[2] We specifically refer to deontological, as opposed to consequentialist, libertarianism.

the neuroscience of decision making is consistent with the standard model of autonomy. However, in one important respect—the role of external influences—the data suggest that at a minimum, it may be wise to reconsider our traditional conception of the subject.

For those steeped in the tradition of experimental science, such empirical data are often viewed as sufficiently compelling to justify the limited claims that we make. For those considering the question from the perspective of other disciplines, the appeal to empirical data may be less persuasive, and our claims may seem bolder than the data appear to allow. Whether scientific investigations can provide a strong foundation for the construction of our ethical theories and their application through social policies is a contentious area within philosophical and political discourse. Elsewhere we have argued that data from the neurosciences *should* be incorporated into our ethical practices (Felsen et al. 2010; Perry and Felsen 2010), but of course the neurosciences do not have hegemony over philosophical thought. Rather, tethering careful consideration of issues such as autonomy to the emerging understanding of how the brain functions provides a singular opportunity to deepen our understanding of the nature of our lived experience as humans.

REFERENCES

Adams, F., and K. Aizawa. 2009. Why the mind is still in the head. In *Cambridge handbook of situated cognition*, ed. P. Robbins and M. Aydede, 78–95. Cambridge: Cambridge University Press.

Alderson, P. 2007. Competent children? Minors' consent to health care treatment and research. *Social Science & Medicine* 65(11): 2272–2283.

Ariely, D. 2008. *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper.

Ariely, D., G. Loewenstein, and D. Prelec. 2003. "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics* 118(1): 73–105.

Armstrong, K., J. S. Schwartz, G. Fitzgerald, et al. 2002. Effect of framing as gain versus loss on understanding and hypothetical treatment choices: Survival and mortality curves. *Medical Decision Making* 22(1): 76–83.

Badre, D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences* 12(5): 193–200.

Badre, D., and M. D'Esposito. 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience* 10(9): 659–669.

Bargh, J. A., M. Chen, and L. Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology* 71(2): 230–244.

Bechara, A., H. Damasio, D. Tranel, et al. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275(5304): 1293–1295.

Becker, G. S. 1976. *The economic approach to human behavior*. Chicago: University of Chicago Press.

Berlin, I., H. Hardy, and I. Harris. 2002. *Liberty: Incorporating four essays on liberty*. New York: Oxford University Press.

Bliss, T. V., and T. Lomo. 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology* 232(2): 331–356.

Braver, T., and J. Cohen. 2000. On the control of control: The role of dopamine in regulating prefrontal function and working memory. In *Control of cognitive processes: Attention and performance* XVIII, ed. Monsell, S. and J. Driver, 713–737. Cambridge: MIT Press.

Buchman, D. Z. and B. J. Russell. 2009. Addictions, autonomy and so much more: A reply to Caplan. *Addiction* 104(6): 1053–1054; author reply 1054–1055.

Caplan, A. 2008. Denying autonomy in order to create it: The paradox of forcing treatment upon addicts. *Addiction* 103(12): 1919–1921.

Chase, V. M., R. Hertwig, and G. Gigerenzer. 1998. Visions of rationality. *Trends in Cognitive Sciences* 2(6): 206–214.

Christman, J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21(1): 1–24.

Christman, J. P. 1989. T*he inner citadel: Essays on individual autonomy*. New York: Oxford University Press, USA.

Christman, J. P., and J. Anderson. 2005. *Autonomy and the challenges to liberalism: New essays*. Cambridge: Cambridge University Press.

Churchland, P. M. 1981. Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78(2): 67–90.

Clark, A. and D. Chalmers 1998. The extended mind. *Analysis* 58(1): 7–19.

Committee on Bioethics. 1995. Informed consent, parental permission, and assent in pediatric practice. *Pediatrics* 95(2): 314–317.

Covey, J. 2007. A meta-analysis of the effects of presenting treatment benefits in different formats. *Medical Decision Making* 27(5): 638–654.

Custers, R., and H. Aarts. 2010. The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science* 329(5987): 47–50.

Damasio, A. R. 1994. *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Damasio, A. R. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences* 351(1346): 1413–1420.

Damasio, A. R., D. Tranel, and H. Damasio. 1990. Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioral & Brain Research* 41(2): 81–94.

De Martino, B., D. Kumaran, B. Seymour, et al. 2006. Frames, biases, and rational decision-making in the human brain. *Science* 313(5787): 684–687.

De Sousa, R. 1990. *The rationality of emotion*. Cambridge, MA: MIT Press.

Dennett, D. C. 1991. *Consciousness explained*. New York: Little, Brown.

Desimone, R. 1996. Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA* 93(24): 13494–13499.

Dunn, B. D., T. Dalgleish, and A. D. Lawrence. 2006. The somatic marker hypothesis: A critical evaluation. *Neuroscience & Biobehavioral Reviews* 30(2): 239–271.

Dworkin, G. 1976. Autonomy and behavior control. *Hastings Center Report* 6(1): 23–28.

Dworkin, G. 1988. *The theory and practice of autonomy.* Cambridge: Cambridge University Press.

Evans, J. S., D. E. Over, and K. I. Manktelow. 1993. Reasoning, decision making and rationality. *Cognition* 49(1–2): 165–187.

Fellows, L. K., and M. J. Farah 2005. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex* 15(1): 58–63.

Felsen, G., L. Whiteley, R. Nadler, et al. 2010. Neuroscience evidence *should* be incorporated into our ethical practices. *AJOB Neuroscience* 1(4): 36–38.

Flanagan, O., H. Sarkissian, and D. Wong. 2007. Naturalizing ethics. In *Moral psychology: The evolution of morality*, ed. W. Sinnott-Armstrong, vol. 1, 1–26. Cambridge, MA: MIT Press.

Frank, R. H. 1988. *Passions within reason: The strategic role of the emotions*. New York, Norton.

Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1): 5–20.

Fuster, J. M. 1995. *Memory in the cerebral cortex*. Cambridge, MA: MIT Press.

Fuster, J. M., and G. E. Alexander. 1973. Firing changes in cells of the nucleus medialis dorsalis associated with delayed response behavior. *Brain Research* 61: 79–91.

Gaylin, W., and B. Jennings. 2003. *The perversion of autonomy: Coercion and constraints in a liberal society*. Washington, DC: Georgetown University Press.

Gillett, G. 2009. Intention, autonomy, and brain events. *Bioethics* 23(6): 330–339.

Gleichgerrcht, E., A. Ibanez, M. Roca, et al. 2010. Decision-making cognition in neurodegenerative diseases. *Nature Reviews Neurology* 6(11): 611–623.

Glimcher, P. W., and A. Rustichini. 2004. Neuroeconomics: The consilience of brain and decision. *Science* 306(5695): 447–452.

Gold, J. I., and M. N. Shadlen. 2007. The neural basis of decision making. *Annual Review of Neuroscience* 30: 535–574.

Goldman-Rakic, P. S. 1995. Cellular basis of working memory. *Neuron* 14(3): 477–485.

Gurmankin, A. D., J. Baron, J. C. Hershey, et al. 2002. The role of physicians' recommendations in medical treatment decisions. *Medical Decision Making* 22(3): 262–271.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4): 814–834.

Harlow, J. M. 1868. Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society* 2: 327–347.

Hazy, T. E., M. J. Frank, and R. C. O'Reilly. 2007. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences* 362(1485): 1601–1613.

Hebb, D. O. 1949. *The organization of behavior*, New York: John Wiley and Sons.

Hill, T. 1989. The Kantian conception of autonomy. In *The inner citadel: Essays on individual autonomy*, ed. J. P. Christman, 91–108. New York: Oxford University Press, USA.

Hofstadter, D. R. 2007. *I am a strange loop*. New York: Basic Books.

Hyman, S. E., R. C. Malenka, and E. J. Nestler. 2006. Neural mechanisms of addiction: The role of reward-related learning and memory. *Annual Review of Neuroscience* 29: 565–598.

Jung, R. E., and R. J. Haier 2007. The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral & Brain Science* 30(2): 135–154; discussion 154–187.

Kable, J. W., and P. W. Glimcher. 2009. The neurobiology of decision: Consensus and controversy. *Neuron* 63(6): 733–745.

Kant, I., P. Guyer, and A. W. Wood. 1999. *Critique of pure reason*. Cambridge: Cambridge University Press.

Kant, I., A. Wood, and J. Schneewind. 2002. *Groundwork for the metaphysics of morals*. New Haven, CT: Yale University Press.

Kuhberger, A. 1998. The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes* 75(1): 23–55.

Kuo, W. J., T. Sjostrom, Y. P. Chen, et al. 2009. Intuition and deliberation: Two systems for strategizing in the brain. *Science* 324(5926): 519–522.

Laughlin, S. 1981. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung C* 36(9–10): 910–912.

Laughlin, S. B., R. R. de Ruyter van Steveninck, and J. C. Anderson. 1998. The metabolic cost of neural information. *Nature Neuroscience* 1(1): 36–41.

Levy, N. 2006. Autonomy and addiction. *Canadian Journal of Philosophy* 36(3): 427–447.

Litt, A., C. Eliasmith, and P. Thagard. 2008. Neural affective decision theory: Choices, brains, and emotions. *Cognitive Systems Research* 9(4): 252–273.

Mackenzie, C., and N. Stoljar. 2000. Introduction: Autonomy refigured. In *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*, 3–31. New York: Oxford University Press.

Marewski, J. N., W. Gaissmaier, and G. Gigerenzer. 2010. Good judgments do not require complex cognition. *Cognitive Processing* 11(2): 103–121.

McClure, S. M., J. Li, and D. Tomlin, et al. 2004. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44(2): 379–87.

McNaughton, B. L., and R. G. M. Morris. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends in Neurosciences* 10(10): 408–415.

McNeely, P. D., P. C. Hebert, R. E. Dales, et al. 1997. Deciding about mechanical ventilation in end-stage chronic obstructive pulmonary disease: How respirologists perceive their role. *Canadian Medical Association Journal* 156(2): 177–183.

McNeil, B. J., S. G. Pauker, H. C. Sox, Jr., et al. 1982. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine* 306(21): 1259–1262.

Miller, E. K., and J. D. Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24: 167–202.

Milner, B. A. 1963. Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology* 9(1): 90–100.

Montague, P. R., and G. S. Berns. 2002. Neural economics and the biological substrates of valuation. *Neuron* 36(2): 265–284.

Montague, P. R., P. Dayan, and T. J. Sejnowski. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience* 16(5): 1936–1947.

Muller, S., and H. Walter. 2010. Reviewing autonomy: Implications of the neurosciences and the free will debate for the principle of respect for the patient's autonomy. *Cambridge Quarterly of Healthcare Ethics* 19(2): 205–217.

Murphy, E. R., J. Illes, and P. B. Reiner. 2008. Neuroethics of neuromarketing. *Journal of Consumer Behavior* 7(4–5): 293–302.

Noggle, R. 2005. Autonomy and the paradox of self creation: Infinite regresses, finite selves, and the limits of authenticity. In *Personal autonomy*, ed. J. S. Taylor, 87–108. New York: Cambridge University Press.

Perry, C., and G. Felsen. 2010. Abortion law should align with evidence from neuroscience. *American Journal of Bioethics* 10(12): 49–51.

Quine, W. V. O. 1969. *Epistemology naturalized. Ontological relativity and other essays.* New York: Cambridge University Press.

Redish, A. D. 2004. Addiction as a computational process gone awry. *Science* 306(5703): 1944–1947.

Redish, A. D., S. Jensen, and A. Johnson. 2008. A unified framework for addiction: Vulnerabilities in the decision process. *Behavioral & Brain Science* 31(4): 415–437; discussion 437–487.

Reiner, P. B. 2010. The rise of neuroessentialism. In *The Oxford handbook of neuroethics*, ed. J. Illes and B. Sahakian, 161–175. Oxford: Oxford University Press.

Romo, R., C. D. Brody, A. Hernandez, et al. 1999. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399(6735): 470–473.

Root-Wolpe, P. 1998. The triumph of autonomy in American bioethics. In *Bioethics and society: Constructing the ethical enterprise*, ed. R. deVries and J. Subedi, 38–59. Upper Saddle River, NJ: Prentice Hall.

Rosch, E. 1999. Reclaiming concepts. *Journal of Consciousness Studies* 11(12): 61–77.

Rupert, R. D. 2009. *Cognitive systems and the extended mind.* New York: Oxford University Press.

Schacter, D. L., G. S. Wig, and W. D. Stevens. 2007. Reductions in cortical activity during priming. *Current Opinion in Neurobiology* 17(2): 171–176.

Schultz, W., P. Dayan, and P. R. Montague. 1997. A neural substrate of prediction and reward. *Science* 275(5306): 1593–1599.

Seymour, B., and S. M. McClure. 2008. Anchors, scales and the relative coding of value in the brain. *Current Opinion in Neurobiology* 18(2): 173–178.

Shiv, B., G. Loewenstein, and A. Bechara. 2005. The dark side of emotion in decision-making: When individuals with decreased emotional reactions make more advantageous decisions. *Brain Research Cognitive Brain Research* 23(1): 85–92.

Sugden, R. 1991. Rational choice: A survey of contributions from economics and philosophy. *Economic Journal* 101(407): 751–785.

Sugrue, L. P., G. S. Corrado, and W. T. Newsome. 2005. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nature Review Neuroscience* 6(5): 363–375.

Suhler, C. L., and P. S. Churchland. 2009. Control: Conscious and otherwise. *Trends in Cognitive Science* 13(8): 341–347.

Sullivan, K. E., P. C. Hebert, J. Logan, et al. 1996. What do physicians tell patients with end-stage COPD about intubation and mechanical ventilation? *Chest* 109(1): 258–264.

Sunstein, C. R., and R. H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70(4): 1159–1202.

Sutton, R. S., and A. G. Barto. 1998. *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Swindell, J. S., A. L. McGuire, and S. D. Halpern. 2010. Beneficent persuasion: Techniques and ethical guidelines to improve patients' decisions. *Annals of Family Medicine* 8(3): 260–264.

Taylor, J. 2005. *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*. Cambridge: Cambridge University Press.

Thaler, R. H., and C. R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT" Yale University Press.

Tom, S. M., C. R. Fox, C. Trepel, et al. 2007. The neural basis of loss aversion in decision-making under risk. *Science* 315(5811): 515–8.

Tremblay, L., and W. Schultz. 1999. Relative reward preference in primate orbitofrontal cortex. *Nature* 398(6729): 704–708.

Trepel, C., C. R. Fox, and R. A. Poldrack. 2005. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Brain Research Cognitive Brain Research* 23(1): 34–50.

Tulving, E., and D. L. Schacter. 1990. Priming and human memory systems. *Science* 247(4940): 301–6.

Turoldo, F. 2009. Responsibility as an ethical framework for public health interventions. *American Journal of Public Health* 99(7): 1197.

Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211(4481): 453–458.

Volkow, N. D., and T. K. Li. 2004. Drug addiction: the neurobiology of behaviour gone awry. *Nature Review Neuroscience* 5(12): 963–970.

Von Neumann, J., O. Morgenstern, H. W. Kuhn, et al. 1947. *Theory of games and economic behavior*. Princeton, NJ, Princeton University Press.

Wallis, J. D., K. C. Anderson, and E. K. Miller. 2001. Single neurons in prefrontal cortex encode abstract rules. *Nature* 411(6840): 953–956.

Watson, G. 1975. Free agency. *Journal of Philosophy* 72(8): 205–220.

Wiggs, C. L., and A. Martin. 1998. Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology* 8(2): 227–233.