

## **From desire to subjective value: what neuroeconomics reveals about naturalism**

DANIEL F. HARTNER

*Rose-Hulman Institute of Technology*

**Abstract:** Philosophers now regularly appeal to data from neuroscience and psychology to settle longstanding disputes between competing philosophical theories, such as theories of moral decision-making and motivation. Such naturalistic projects typically aim to promote continuity between philosophy and the sciences by attending to the empirical constraints that the sciences impose on conceptual disputes in philosophy. This practice of checking philosophical theories of moral agency against the available empirical data is generally encouraging, yet it can leave unexamined crucial empirical assumptions that lie at the foundations of the traditional philosophical disputes. To illustrate this, I compare recent work in the neuroscience of decision to traditional philosophical theories of motivation and argue that the traditional theories are largely incompatible with empirical developments. This shows that genuine continuity between philosophy and science means that in some instances the conceptual foundations required to explain the phenomenon of interest be developed by the sciences themselves.

**Keywords:** naturalism, neuroeconomics, decision-making, expected utility theory, folk psychology, moral motivation, neurophilosophy

**JEL Classification:** B40, B41, D87

[I]t is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking (Anscombe 1958, 1).

A great deal of the recent work in cognitive science has, tacitly or explicitly, assumed very much the picture of mental organization that folk psychology proposes. There are other straws in the wind, however. There are findings and theories suggesting that something is seriously wrong with the simple belief-desire structure implicit in common sense wisdom (Stich 1983, 230).

---

**AUTHORS' NOTE:** I am grateful to John Bickle, Thomas Wells, and an anonymous referee for many helpful comments and suggestions for revision. I would also like to thank Tim Schroeder, Dimitria Gatzia, and Gonzalo Munevar for their helpful comments on, and criticisms of, earlier versions of this paper.

The last decade or so has seen a growing number of philosophers express concern over the proliferation of dubious empirical claims and assumptions in ethics. Stephen Darwall, Alan Gibbard, and Peter Railton, in their overview of the last century of work in ethics, observe that, “too many moral philosophers [...] have been content to invent their psychology or anthropology from scratch” (Darwall, et al. 1992, 188-189). John Doris and Stephen Stich have echoed that concern, arguing further that philosophy’s empirical complacency has discouraged scientists from “undertaking philosophically informed research on ethical issues” (2007, 115).

For those of us who share these concerns about philosophy’s empirical commitments in an age of rapid scientific progress, it might seem encouraging that there is now at least one problem in metaethics and philosophical moral psychology that is receiving extensive empirical treatment from philosophers. That is the problem of moral motivation (MM), i.e., the problem of explaining, perhaps conceptually, the nature of the relationship between an agent’s moral judgments (or beliefs) and her behavior.

The conceptual difficulty that lies at the heart of the problem of MM is straightforward. Suppose that I come to believe that the morally right thing to do is to tithe my salary in support of famine relief. Does my believing this mean that I will necessarily be motivated to do it? On the one hand, it seems so because if I should insist that giving is the right thing to do without actually being so motivated, the best explanation for my lack of motivation might be that I do not genuinely believe what I claim to. On the other hand, it seems plausible that I might genuinely believe that I should tithe my salary and yet remain unmotivated precisely because I do not actually want to. Which is the better account of the relationship between my moral judgment and my motivation? In very plain terms, this is the problem of MM.

In its more rigorous academic form, the problem of MM encompasses at least two distinct though related philosophical disputes concerned with whether and how moral judgments motivate moral agents. The first dispute is about *whether* moral judgments motivate. Motivational internalists argue that moral beliefs motivate *necessarily* while externalists deny this. The second dispute is about *how* such judgments motivate. Proponents of the so-called Humean theory of motivation (or Humeanism) argue that moral beliefs are *insufficient*

for motivating agents since motivation requires in addition to a belief the presence of a conative state such as a desire. Anti-Humeans reject the Humean theory on the grounds that moral beliefs are themselves sufficient for motivation. Some anti-Humeans endorse internalism, or one of a few related ideas such as that moral beliefs are somehow simultaneously desire-like (“besires”)<sup>1</sup> or that moral beliefs co-occur with or otherwise trigger the relevant desires.

The connection between this philosophical dispute and empirical science is straightforward. Scientists too are interested in the relationship between value judgment, decision-making, and motivation. It might seem promising, then, that a growing number of philosophers—naturalists, though the label is perhaps not always self-applied—are using data from psychology, psychiatry, cognitive science, and neuroscience to help resolve these longstanding philosophical disputes about MM. Naturalism is the philosophical position that the aims and methods of philosophy are continuous with those of the empirical sciences.<sup>2</sup> Different philosophers define naturalism in different ways but, at least in practice, most naturalists are committed to the idea that the sciences are continuous with philosophy in the sense that the empirical facts ought to constrain and inform the development of or choice between philosophical theories in some way, particularly when those theories purport to explain phenomena that are of interest to the sciences. So, in the case of MM, naturalists try to show that the data vindicates one or another of these traditional theories. The empirical data is a good place to look when purely conceptual considerations fail to settle the matter.

For example, Roskies (2003, 2006) argues that patients suffering from damage to the ventromedial prefrontal cortex (VMPFC) serve as counterexamples to internalism.<sup>3</sup> Kennett and Fine (2009) have argued that clinical research with autistic patients and psychopaths supports a Kantian account of motivation, or some form of anti-Humeanism according to which moral judgments are necessarily motivating.

---

<sup>1</sup> The term, coined by Altham (1986), is now commonplace in the relevant metaethics literature.

<sup>2</sup> This way of putting it is commonly attributed to Quine (1969), who made a similar point about psychology and epistemology, though few philosophers today endorse Quine’s account of the relationship between philosophy and the sciences.

<sup>3</sup> More recently, Schroeder, Roskies, and Nichols, have argued that instrumentalism—a variation on the Humean theory, which holds that an agent is motivated when she forms beliefs about how to satisfy her pre-existing desires—“fits well with the neuroscientific picture” of motivational processes (2010, 106).

Prinz (2006) uses data on psychopathology, though, to argue for a Humean sentimentalist account of moral concepts according to which an agent's believing that an action is morally wrong amounts to her having a sentiment of disapprobation toward it. These are just some of many available examples of empirically sophisticated naturalism in contemporary moral philosophy.

My goal in this paper is to raise the concern that many such naturalistic projects, despite their empirical sophistication, share a problematic core commitment. These philosophers are quite right to recognize the limits of traditional philosophical methods like conceptual analysis, intuition, and armchair reflection for elucidating the nature of MM. They endorse the rather plausible idea that scientific data has much to offer these conceptual inquiries into judgment, decision, and motivation. But there is another idea at work in each of these approaches that is, I think, considerably less plausible. That is the idea that scientific research will ultimately preserve or prove sufficiently compatible with the framework of commonsense psychology in which philosophical disputes about MM are couched.

There is a kind of realism about folk psychological (FP) concepts like belief and desire involved in philosophical disputes about moral motivation. Humeans, anti-Humeans, internalists, and externalists are all, in some sense, in dispute about the role that these states play in bringing about MM. They take for granted that the right or best account of the relationship between moral judgments and motivation will preserve beliefs and desires (or something near enough). After all, this is the very point of using data to vindicate one or another of these traditional theories. To put it another way, eliminativism, instrumentalism, and other forms of serious skepticism about FP states are neither forms of anti-Humeanism nor externalism—such views avoid this FP framework altogether.<sup>4</sup>

The last ten years or so of work on the neuroscience of value judgment, decision-making, and motivation has produced a reasonably unified field called neuroeconomics. Work in neuroeconomics ranges

---

<sup>4</sup> Eliminativism is not a form of anti-Humeanism because the latter theory holds not just that beliefs are insufficient for motivation—a claim that might seem compatible with the nonexistence of FP states—but also that motivation requires the presence of a *desire* (or related FP state). Eliminativism is not a form of externalism because it seems there is not much sense in the eliminativist's taking a specific position on the effects of undergoing nonexistent states. Stich has made a similar point in response to Dennett's instrumentalism, arguing, for example, that only real entities and not useful fictions can have causes and effects (1983, 244).

from cellular-level to social-level neuroscience and it has been converging on an account of the causal mechanisms of value judgment and motivation. This account—which is, I think, immediately relevant to the issue of moral and social cognition and hence to philosophical disputes about MM—neither invokes commonsense FP states directly nor appears likely to lend itself to accurate redescription in FP terms.

I will proceed by outlining the development of neuroeconomics and its core concepts, like subjective value, that are important for connecting this scientific field to philosophical disputes about MM. I hope that readers will forgive the empirical review that occupies the first part of this paper. It is crucial to understanding the conceptual connection—one that I think we can no longer afford to ignore—between neuroeconomics and philosophical moral psychology. I then argue that neuroeconomics is poised to raise two serious challenges for traditional philosophical theories of moral agency which rest upon the outmoded framework of FP. Finally, I will argue that what we should conclude from this is not that neuroeconomics has closed the case on explaining how decision-making, value judgment, and MM work, but rather that the relationship between a science like neuroeconomics and philosophical moral psychology can teach us a good deal about what philosophical naturalism properly amounts to, that is, about what it means to develop genuine continuity between philosophy and the sciences.

This last point about naturalism is the real heart of this paper. In their work naturalists tacitly and sometimes even explicitly reveal a commitment to the idea that philosophy is prior to science in at least one important way: carving out and explicating the concepts that scientists require is in large part philosophical, not scientific, work. For example, Jackson and Pettit (1990) are among those naturalists who have been explicit about this idea. They argue that at the very least, a “completed neuroscience” will have to reveal that the folk roles (which is their term for commonsense functional roles) of beliefs and desires are in fact occupied, since with just a bit of conceptual analysis we can see it is sufficient for having beliefs and desires that the folk roles be occupied (1990, 36). It is philosophy, not the sciences, that reveals this.

In philosophical disputes about MM, it is FP concepts that guide discussion. After all, the problem arises from a conceptual difficulty: the question is *whether* motivation is a necessary component of genuine belief, where motivation is an FP concept already widely analyzed and

cemented in the philosophical literature. Then, where the conceptual contributions run dry, naturalists turn to empirical data to tip the scales in favor of one philosophical theory or another. But what the case of neuroeconomics and MM shows is that there may be good reason to think that the sciences themselves carve out and explicate the concepts they require as they develop. Thus, the argument I give here against FP is not intended as an argument for eliminativism but rather as an illustration of how the sciences, in this case economic theory, can lead philosophy in its conceptual development—a twist that many philosophers and even self-described naturalists may find *prima facie* implausible. In fact, though, it is the philosophy-first form of naturalism, which forces empirical data to fit with traditional philosophical theories, that has in many cases become a barrier to recognizing and developing genuine continuity between philosophy and science.

#### **VALUE AND CHOICE: SOME DEVELOPMENTS IN NEUROECONOMICS**

Two different academic camps continue to contribute to the conceptual and empirical literature on human and primate decision-making and judgments of economic value. Traditionally, philosophers seek to clarify the relevant psychological concepts and distinctions used in explanations of cognition and behavior while cognitive scientists and neuroscientists collect data to elucidate the physiological mechanisms underlying (or constraining) these explanations. On this way of distinguishing these two camps, the division of labor is straightforward, as is the potential for collaboration between the two camps. We should expect, if this division largely holds true, that philosophers might take an interest in the ways in which their theories and concepts can be mapped onto data from neurophysiology. Conversely, scientists might sometimes invoke or borrow philosophical concepts and distinctions to enrich their more mechanical explanations. Indeed, as I argued in the previous section, the former kind of interaction is precisely what we have seen, with self-described naturalists increasingly looking for ways to map their concepts onto neurological explanations. But developments in decision science have begun to complicate this relationship.

The last decade has seen the development of neuroeconomics, which weds behavioral economics with experimental neuroscience. Its key methodological innovation is to use well-vetted theories in economics, like expected utility theory (EUT), to contextualize neural data generated

by subjects engaged in tasks of judgment and decision. To borrow a helpful summary from one of the field's founders, Paul Glimcher (2009), the idea is that behavioral economists can use neuroscience to reveal the physiological constraints on real agents that sometimes lead them to violate the axioms of normative economic models like EUT, while neuroscientists can use economic theory to help develop algorithmic models of decision-making for which they can identify relevant neural mechanisms (Glimcher, et al. 2009, 7-8).

In what follows I shall assume that neuroeconomics, which seeks to explicate decision-making, is relevant to disputes about *moral* decision-making. In lieu of a lengthy defense of this idea, I offer just two brief remarks here. First, the idea that moral judgment and moral decision-making is a species of value judgment and decision-making more generally is rather plausible on its face. Indeed the idea must also be plausible to philosophers who engage in philosophical moral psychology, since in that field's literature the relevant FP states like belief and desire are the very same as those that appear in (non-moral) philosophy of mind and philosophical psychology. To reject the relevance of neuroeconomics to moral philosophy on the general grounds that economic decision-making should be distinguished from ethics is to reject the idea that moral cognition is a subset of cognition in general. This is not merely a strange idea, but one that seems to be incompatible with current practices in philosophical moral psychology.

Secondly, there are already empirical links in place between decision-making and moral decision-making. The connection comes primarily by way of social neuroscience, a branch of neuroscience concerned with the relationship between patterns of neural activity—most often investigated using functional magnetic resonance imaging (fMRI)—and tasks involving social contexts and norms. Some researchers study the interaction between economic decision-making and moral/social decision-making directly by using or designing tasks that force subjects to make economic decisions in conjunction with decisions about moral and social norms. The idea is to study changes in neural activation patterns as subjects engage in a variety of economic tasks in which social or moral norms are salient. For example, many experiments in social neuroscience use gambling games to force subjects to make decisions about whether and to what extent to cooperate with others in the pursuit of profit (for a discussion of some of these tasks, see Sanfey, et al. 2014). This is one of the

more obvious instances of overlap between economic decision-making and moral/social decision-making. Some economic decisions are simultaneously decisions involving the interests of others.

Moreover, some experiments in social neuroscience have investigated the relationship between social and moral judgments directly. For example, there is some evidence based on patterns of neural activation that judgments about moral norms are a specialized form or subset of judgments about more general (nonmoral) social norms (see Moll, et al. 2002). Insofar as we accept this idea, that moral judgments are *decisions* about what it is right, or best, to do under such-and-such circumstances, then we shall see that neuroeconomics is giving us reason to doubt the empirical adequacy of philosophical disputes about value judgment and motivation couched in FP terms. With this out in the open, I turn to the details.

One primary goal of neuroeconomics—as has been stated explicitly by neuroeconomists—is to establish connections between variables like utility, derived from observable behavior, and “psychophysiological quantities”, like the firing rate of a neuron (Rustichini 2009, 34). In one sense, then, what neuroeconomists are trying to do is to develop a more powerful account of decision than is possible using traditional theories like EUT alone. If we can understand the physiological constraints on real agents that EUT does not take into account, then we will be better able to predict and explain why agents who have to make decisions sometimes violate the axioms of theories like EUT.

EUT uses psychological constructs like utility to make sense of decision-making.<sup>5</sup> Utility is the calculated psychological value of an option that an agent chooses. In economic theory it is defined in relation to rational choice, that is, choice which realizes the greatest possible subjective value for an agent. Rational choice is facilitated by preference orderings over states of the world which reflect an agent’s relative prioritization of those states. Agents choose rationally when they realize their most valued possible preferences given their option set. In this sense it is a kind of psychological concept, but it is important to note that it can only be calculated from behavior—what real agents choose. The problems with EUT as a realistic model of human behavior are well rehearsed (Kahneman and Tversky 1979). Real agents do not always

---

<sup>5</sup> It is worth noting here that the idea that the utility in EUT is a psychological construct is controversial, since some economists understand utility instrumentally, as an index of preferences as patterns of revealed choice. This is a point that I will return to below in discussing conceptual progress in the special sciences.



choose as though they were trying to maximize their utility; they can be induced to violate the axioms of EUT fairly easily. So we might rightly wonder, why bother with EUT at all?

The value to neuroscientists of an economic theory like EUT is that, despite its imperfections, when agents *do* choose in accordance with its axioms—completeness, transitivity and independence—those agents behave *as if* they were trying to maximize their utility. This affords scientists the opportunity to investigate the neurological constraints on agents engaged in value judgments. The insight from neuroeconomics is that it may be possible to find neural mechanisms that fit with EUT when EUT is accurate in predicting choice, and then to narrow the investigation to those mechanisms where EUT goes astray, i.e., when agents violate its axioms, so that the neural data can be used to predict and explain those violations. The result would be something more powerful than EUT because it would capture the physiological constraints that make people, to put it controversially, ‘irrational’. In other words, it becomes possible to accommodate criticisms of EUT, including the famous Allais paradox which shows that people are in fact not ideal choosers since they can be induced to violate at least one axiom of EUT.<sup>6</sup>

The way neuroeconomists have so far gone about doing this is as follows. They begin with a theoretical construct that would allow this neural investigation to unfold. That construct is subjective value (SV). Neuroeconomists have developed a working definition of SV, one that would allow it to harness the power of EUT without the shortcomings. By definition, SVs are the mean firing rates in action potentials per second of specific populations of neurons which predict choices of agents (though stochastically). When expected utilities predict choice, SVs are linearly proportional to the expected utilities. This way the SVs are always consistent with choice (though, again, stochastically) even when choice is not consistent with EUT (Glimcher 2009). SV is thus like utility in that neural activity would track choice in cases in which subjects do choose as though they were maximizing their wellbeing. But SV would deviate from utility and continue to track choice in the cases in which agents violate the axioms of EUT. So it could still be used

---

<sup>6</sup> In the case of the Allais paradox, people’s actual choices between lotteries are influenced by the addition of outcomes that are irrelevant to their relative utility. The axiom violated is the independence axiom. The details are somewhat technical, but the axiom is important to the idea that subjects have well-defined preferences, and it is for this reason that the Allais paradox presents a serious problem for EUT.

to predict choice when EUT fails, as it does in the case of the Allais paradox.

One of the interesting features of SV—and an improvement over the concept of utility—is that the construct accords well with a two-stage model of decision making that has been under development in the neural sciences for several years. The two stages in this model are *valuation* and *choice*. In valuation, subjects assign values or utilities to individual goods or actions in their range of options. At the behavioral level, these economic values are calculated by quantifying the subject's choices relative to the alternatives.<sup>7</sup> At the neural level, these values would need to occupy the role of SV, the theoretical construct that is now central to neuroeconomics for the reasons just given.

The pressing empirical question in neuroeconomics has been whether SV exists—whether there exists a neural firing rate pattern or blood-oxygen-level dependent (BOLD) signal<sup>8</sup> linearly correlated with utility *when* in fact utility does predict choice, and which regions or cell populations of the mammalian brain are capable of encoding this signal. The rather surprising answer emerging from the empirical literature has been that there are cell populations that encode SV, and that the SVs of items (or, in economic terms, goods) are likely encoded in the ventromedial prefrontal cortex (VMPFC) and the SVs of actions likely encoded in the striatum. It is worth briefly reviewing a bit of that evidence.

Recordings from cells in the VMPFC have contributed to the localization of valuation. In a series of studies, Padoa-Schioppa and Assad (2006) presented monkeys with choices between different types and quantities of juices and foods. They then calculated a behavioral-level subjective value for each of the juices based on the monkeys' choices and the quantities offered. They then checked for neuronal activity that might support the behavioral-level subjective value calculations, the hypothesized common currency for choice.

---

<sup>7</sup> For example, if a monkey chooses reward 1A (e.g., one apple slice) when paired with one 1B (e.g., one raisin), 2B (e.g., two raisins), and 3B; it is indifferent at a ratio of 1A:4B; and it chooses B when 6B and 10B are offered, then the value of 1A is roughly equal to the value of 4B [i.e.,  $V(1A) = V(4B)$ ] and hence has a subjective value of approximately 4.

<sup>8</sup> BOLD-contrast imaging is a method used by neuroscientists in functional magnetic resonance imaging (fMRI). The images of brain activity produced by this method are based on changes in the level of oxygen in different areas of the brain associated with changes in levels of activity.

The researchers identified three distinct neuronal patterns corresponding to three types of neuronal function. A portion of the 931 cells in the orbitofrontal cortex (OFC) from which recordings were taken showed a firing rate significantly correlated with the subjective values previously hypothesized from the behavioral data. These neurons were termed *offer value* neurons because they track the subjective value of the juice option offered. In other words, the activity of these particular neurons co-varied with the value of the juice on offer.

A second subset of the neurons showed a firing rate linearly correlated with the subjective value of the juice (i.e., reward) that the monkey *actually* chose (or would eventually choose). In this case the neuronal activity was low when the monkey chose the juice with a chosen value score of about 2, higher when it chose a juice with a value score of about 4, and highest when the monkey chose a juice with a value score of about 6. That these variations in cell activity are significantly correlated seems to indicate that they represent the subjective value of the chosen reward. This subset of neurons was therefore labeled *chosen value* neurons.

The third subset of neurons showed a distinct categorical or binary firing activity response to *particular* juices. The researchers accordingly labeled these *taste* neurons.

Activity in each of these three types of neurons showed a distinct timing pattern. Offer value and chosen value neurons predominantly fired immediately following the presentation of juice options, while taste neurons fired after the juice reward was presented (Padoa-Schioppa and Assad 2006; Kable and Glimcher 2009). Similarly, these three types of neurons have been found in the caudate and putamen of the striatum where research indicates they track the values of actions rather than goods (Samejima, et al. 2005).

Importantly, these studies also showed that the neuronal value representations were menu-invariant. That is, the neural responses are representations of the direct value of individual goods/items rather than representations of their relative value, or the value of a good relative to its paired alternative. Recording activity in 557 individual neurons in the OFC, Padoa-Schioppa and Assad presented monkeys with competing juice pairs (i.e., offers). To determine whether the neuronal responses depended upon the menu (i.e., upon what alternatives were available at that particular time), they recorded the neuronal activity while the monkeys chose between *three* different juices (A, B, and C in

decreasing order of preference) in varying amounts, presented in interleaved pairings of A:B, B:C, and C:A. The results again showed three patterns of neuronal activity corresponding to the three types of neurons (offer value, chosen value, and taste neurons). And these neuronal responses were invariant to changes of menu: the neuronal activity encoding the value of each of the juices was largely independent of availability of other juices.

This is especially important because transitivity of choice at the behavioral level is already well established. Transitivity, which is one of the axioms of EUT, is the basic economic idea that a subject who prefers A to B and B to C must prefer A to C. In behavioral experiments, the monkeys' choices do in fact exhibit transitivity. Monkeys who prefer juice A to juice B and juice B to juice C prefer A to C (Padoa-Schioppa and Assad 2008; Kable and Glimcher 2009). Establishing the menu invariance of neuronal activity is crucial because it shows that the neuronal responses, like the behavioral responses, are stable and consistent, and therefore reflect transitivity (Padoa-Schioppa and Assad 2008). Thus, evidence of transitivity in neuronal activity supports the idea that the values of goods are represented in a common, comparable currency in the OFC neurons. In other words, transitivity is only possible if the neurons encode individual subjective values of goods on a single, common scale and not merely relative (menu-variant) values. Each good on offer, then, has its own *absolute* subjective value represented by particular neurons on a common scale. This, in sum, is the valuation stage in the emerging two-stage model of decision.

In the interest of providing a more complete story of decision with clear relevance to explanations of judgment and motivation, I need to say just a bit about the second stage in the two-stage neural model: choice. It is the valuation stage that is my primary concern here, however the choice stage is important because it provides the overt link to behavior, and hence the relevance of neuroeconomics to disputes about the relationship between judgment or decision and motivation.

The research on choice implicates the lateral prefrontal and parietal cortex. Much of this research is based upon work with the visuo-saccadic control system in the primate brain. Saccades are rapid eye movements executed for the purpose of fixing one's gaze on a scene. Rapid eye movements to points of interest in the visual field help an animal to build a map of a scene. This eye activity is initiated by the visuo-saccadic control system, which includes the frontal eye fields (FEF)

in the cortex and the superior colliculus (SC) in the midbrain. Neuroscientists interested in sensory-motor control have studied this system extensively. It appears to provide the mechanism by which information concerning the chosen option, and not the unchosen options, is implemented in motor systems downstream from the valuation circuitry. It is here that we find the explicit link between decision and motivation, a link crucial to the philosophical account of MM under consideration in this paper.

The details are complex, but the basic idea is that neurons in the lateral intraparietal area (LIP), FEF, and SC form a network for visuo-saccadic decision-making. Studies with monkeys on saccadic decision-making tasks have repeatedly shown that the firing rates of neurons in LIP and FEF increase as evidence accumulates that a visual response will result in reward. Interestingly, once those firing rates cross a preset threshold, a saccade is initiated (Shadlen and Newsome 2001). Further research has since indicated that this firing rate threshold represents a value threshold for movement selection (Roitman and Shadlen 2002). This is the overt link to behavioral output.

It is also worth noting that in the last ten years or so of work in neuroeconomics, much progress has been made on revealing the mechanisms through which SVs—the currency for choice—are learned and represented in the primate brain. Dopaminergic (DA) neurons in the midbrain encode a reward-prediction error (RPE), i.e., the difference between the outcome of an action actually experienced and the predicted outcome of the action (Schultz, et al. 1997). Research indicates that the firing rates of these DA neurons are linearly related to RPE as calculated by behavioral-level economic models (Bayer and Glimcher 2005). Beyond having evidence for the existence of SV as a real neural entity, scientists now have some idea about the neural mechanisms by which these values are learned and encoded in the mammalian brain.

And, as I mentioned briefly above, as these lines of research elucidate the mechanisms behind choice in the primate brain, social and cognitive neuroscientists are revealing that the same regions, most notably the striatum and VMPFC are consistently implicated in tasks in which subjects are asked to make moral and social judgments (e.g., Greene and Haidt 2002; Moll, et al. 2002). While much work remains to be done, there are already some direct links between moral judgment and decision-making and the neurophysiology of decision and motivation.

In sum, both the primary goal and empirical development of neuroeconomics are well established. Neuroeconomics seeks to provide neurocognitive explanations of value judgment and choice behavior, and it has made considerable empirical progress toward that goal by wedding techniques in neuroscience with the theoretical framework of EUT. The overlap with theories in philosophical psychology and moral psychology is also clear: both aim to provide an empirically adequate account of the relationship between value judgment, choice behavior, and motivation. This raises an important challenge for naturalists who want to apply empirical data to longstanding philosophical disputes about judgment and motivation.

### THE CASE AGAINST FOLK PSYCHOLOGICAL EXPLANATIONS

Philosophical theories that employ FP concepts to deal with the phenomena of judgment and motivation, especially those that aim to achieve results compatible with the results in the sciences, will need to show that FP is up to the task of capturing these details. It is here that problems arise since FP theories of motivation ultimately run up against a difficult tradeoff. I will try to show that as FP theories become nuanced enough to track the kinds of explanations that neuroscientists have been developing on the back of years of work in behavioral economics, those theories will tend to require concepts and theoretical postulates that lack the commonsense features that are characteristic of FP's mental state postulates. In particular, they will jeopardize the characteristics of postulates like beliefs and desires that make such states commonsensical or folksy. But, perhaps not surprisingly, to the extent that FP-based theories like Humeanism, anti-Humeanism, internalism and externalism preserve their postulates in commonsensical form, and thus remain true to the folksiness of FP, they will be forced toward a level of generality that is far too coarse to say much of substance about the relationship between value judgment and motivation.

In general terms, FP theories need to map the cognitive-level FP story about an agent's subjectively valuing an item or action onto the neurophysiological mechanisms—subjective value—upon which that story must supervene according to developments in neuroscience. The difficulty for the proponents of FP theories is that (1) SVs “exist”—they are genuine *neural* entities, and (2) their contribution to decision and motivational processes—i.e., their explanatorily relevant characteristics

and functions—pertain exclusively to the biophysical level. Importantly, though, those characteristics and functions were uncovered *not* on the back of commonsense psychology but rather on the back of economic theory. This means that the way that we come to understand the psychological-level contributions to accounts of choice behavior will be guided by theoretical and conceptual development in economic theory, not commonsense psychology. Thus, it can hardly be surprising that FP explanations of decision and motivation find themselves forced to choose between empirical inadequacy and trivial generality. And that is precisely what seems to happen.

Consider what a proponent of FP explanations might say about how states like belief and desire fit into the neurological account described above. For example, one possibility for the proponent of FP would be to insist that the power of FP lies in its generality. We vindicate FP when we simply link or identify a state like desire with subjective value. So we might say that choice involves selecting from among objects for which we feel competing levels of desire. A monkey faced with a choice between grapes, bananas, and raisins is essentially faced with the task of selecting from among competing desires for each of the fruits, and perhaps chooses on the basis of beliefs about the quantities available. Two grapes, the monkey *believes*, satisfy its *desires* better than one raisin. Dopamine, synaptic plasticity, learning, and so on, are merely the lower-level neurophysiological mechanisms upon which the cognitive events must supervene given the inevitability of beliefs and desires (see Jackson and Pettit 1990).

The problem is that this approach seems to jeopardize the causal relevance of beliefs and desires as understood in the going philosophical theories of MM. For example, Humeans claim that moral beliefs are insufficient for motivation because beliefs require the presence (or co-occurrence) of a desire to motivate. Anti-Humeans deny this, generally because they are drawn to some kind of motivational internalism. On the account just given, the Humean theory is—on the most charitable reading—just trivially true. It is true in a manner of speaking that desires are required for motivation. But the requirement is trivial, failing to provide a meaningful explanation of the target phenomenon, because desires in this sense are present to varying degrees in *all* of the options, including those that are ultimately bypassed by the chooser. It is this latter point that Humeanism simply overlooks. The addition of desire cannot be an adequate explanation for

motivation if varying degrees of desire are already in place in each of the options. Desire, understood as SV, must be present for motivation precisely because in any real choice SV is always present. Thus, even on a charitable reading, Humeanism as a way explaining the phenomenon is true but nevertheless very far from insightful or explanatory. There is much more going on here than the mere presence or co-occurrence of desire in the FP sense. Proponents of FP theories will need a better way to defend the explanatory relevance of their (disputes about) FP mental state postulates in light of the data.

Another, perhaps more plausible, proposal to suggest on behalf of FP is that desires are somehow linked to utilities rather than SVs. After all, utilities in decision theory are psychological constructs about the value of an item to the agent, which more or less amounts to the commonsensical idea of desirability (I will return to this point again in the final section). Moreover, since desires like utilities sometimes fail to predict choice, this seems an especially appealing possibility.

On this proposal, however, we end up facing precisely the same difficulty that has hampered EUT since the Allais paradox. Utilities are regarded as psychological entities, but as an empirical fact they are revealed behaviorally, by physical choice selection. This means that, in identifying desires with utilities, it will have to be true that utilities, and hence desires, always predict choice. But the Allais paradox shows that they do not. More problematic still, one need not even appeal to technical developments in economics to make this point. As Gauker (2005) points out in addressing the conceptual matter of the relationship between beliefs and desires (i.e., the possibility of a so-called belief-desire law that relates beliefs to desires conceptually), the simple fact is that people do not always do what they most desire to do. Any theory that requires this is already false. Utility simply is not a perfect predictor of choice.

So far the problem encountered is one of explanatory adequacy: because of the simplicity and generality of FP, FP theories of decision and motivation lack explanatory power. One solution is to add a bit of complexity to our FP concepts so as to permit FP explanations to keep pace with the advances in the science of decision. But as we work to find ways to make FP fit with the developments in neuroeconomics, we have to keep in mind that for FP to *be* FP, we need to preserve the basic, commonsense ideas about what these states amount to. The folksiness



of FP is an important constraint on the amount of complexity we can add to FP explanations.

Traditionally, folksiness has meant understanding beliefs and desires in terms of directions of fit. The basic idea behind directions of fit is usually attributed to Anscombe (1957), but it is now widely discussed in the philosophical literature. One formulation, from Schroeder (2009), holds that desires have a world-to-mind direction of fit, which makes them like imperatives that are satisfied when the world changes as they command. Beliefs, by contrast, have a mind-to-world direction of fit, and so are like declarative sentences, satisfied when they are made to match the world. But as we look for ways to map these concepts onto neural explanations, adding nuance to gain explanatory adequacy, we are forced away from these commonsensical characteristics of FP states.

The idea of directions of fit is useful (perhaps as a metaphor) for explaining or rationalizing our neat, intuitive conceptual distinction between commonsense notions of belief and desire, but, as we have just seen, we need something more nuanced than directions of fit to adequately explain why agents choose what they do. As the Allais paradox and Gauker's argument about a belief-desire law show, people do not in fact simply go around shaping the world to their minds. Such an explanation for choice behavior is not just unsatisfying but false. If the commonsense concept of desire can be made relevant to the explanation of choice, it will not be by virtue of a vague (or possibly metaphorical) explanation of commonsense psychological states in terms of directions of fit. For this would require an explanation of choice in terms of desire-as-world-shaping that explains why a state that is *defined* in terms of an agent's world shaping is a state that often flatly fails to predict how an agent will in fact attempt to shape the world. The point is that this shortcoming can be corrected, of course, by supplementing the FP definition with a list of caveats that help us to better predict and explain choice behavior. But there is nothing commonsensical or folksy about mental state concepts loaded with asterisks explaining the various exceptions. More importantly still, the very act of developing that list of exceptions looks much more like the undertaking of an empirical investigation than the application of an indispensable postulate of a folk theory.

But there is perhaps another route for the proponent of FP to try. Suppose, again, that we insist on the connection between FP postulates

like beliefs and desires and the psychological construct of utility. For the reasons we have just seen, equating desires with either SVs or utilities will not work. But we might instead argue that an agent has representations of facts about the value or worth of each of their options. That is, the agent has something more like *beliefs* in the FP sense about the values of the options on offer. This avoids the previously discussed problems with saying that desires are like utilities, while still preserving a crucial role for FP postulates. Now the utilities are linked to beliefs—beliefs about the worth of the items on offer. Yet this seems to get the traditional dispute about Humeanism backwards. In the traditional dispute, the question was not how beliefs about values tip the scales in cases of competing desires but precisely the opposite. Humeans claim that we need desires to tip the scales in motivating us to act in accordance with our beliefs. So such an account might find some room for FP but only by turning the dispute about Humeanism in the wrong direction. Such a result would upend any claim about the indispensability, accuracy, or utility of FP.

These proposals no doubt fail to exhaust the possible FP-friendly interpretations of data from neuroscience. However, that is really not my goal here. Rather, this discussion is illustrative. In considering just a few of the most readily apparent ways to preserve FP explanations a difficult tradeoff already emerges. As neuroscience advances, theories rooted in FP are likely to face a difficult choice between advancing their empirical adequacy and maintaining conceptual coherence. More importantly, though, it seems to me that this result is to be expected given the way in which the concepts most relevant to the explanation of choice and motivation have developed along with the sciences.

### CONCEPTUAL DEVELOPMENT IN THE SPECIAL SCIENCES

The trouble with the traditional philosophical accounts of moral judgment and motivation is that they rest on folk psychology, while the neuroscience of decision has scaled down to lower levels of investigation such as brain regions and cells, not from FP but from economic theory, where the relevant psychological concepts emerge from behavioral data not preliminary philosophical hunches formed independently of any systematic empirical investigation into how best to distinguish mental state types. This is precisely the sense in which conceptual progress is guided by the sciences themselves, in this case by economic theory, rather than prior commitments to purportedly

indispensable conceptual frameworks. This in turn grounds my claim that putatively naturalistic attempts to interpret the data to fit traditional philosophical frameworks may in fact run contrary to naturalism in the sense that they implicitly promote the traditional and artificial demarcation between *a priori* philosophy of mind and empirical science.

On the traditional division of labor sketched earlier, conceptual developments and clarifications are expected to set the constraints on scientific theorizing. This is precisely why proponents of FP are inclined to find ways to reconcile the concepts of FP with those of neuroeconomics. These theories must be congruous, the argument goes, because concepts like subjective value in neuroeconomics are products of basic conceptual clarifications: something must occupy the role of desire in economic theory. Jackson and Pettit (1990), as noted above, have articulated this argument explicitly. Yet the results of the preceding section show that something has gone wrong with this argument.

Firstly, it seems that no conceptual role in neuroeconomics is straightforwardly consistent with FP's conceptual apparatus, or with acceptable modifications of it. There seems no clear place for desire in the traditional FP sense in neuroeconomics. At least, there is presently no clear way of defending the idea that FP concepts add anything of importance (let alone that they are indispensable) to the explanatory account of decision-making under development.

Secondly, and more importantly, even if some FP concept could occupy a role in neuroeconomics, the fact is that the conceptual role was nevertheless *not* a product of prior philosophy or commonsense psychology but rather a product of the development of economic theory. The postulates from which neuroeconomists are working are those of such economic theories as EUT, not commonsense psychology. This is because it is EUT rather than FP that furnishes the concepts, like utility, that are based on empirical measurements (or experiments designed to take those measurements) that must be accounted for in the final explanation. Utility in EUT is a way of characterizing raw data, i.e., a set of observations about the empirical world—observable and quantifiable selections or preferences of individual agents—that stands in need of explanation. The goal of developing an explanation of decision-making in neural terms is to explain precisely that set of raw data about choices or preferences. It is no surprise then that the

concepts needed to quantify those preferences will serve as the conceptual foundations of the newly developed theory. It is utility, not desire, which describes those observable choice selections or, in psychological parlance, preferences. If I choose one grape to one apple, I exhibit (behaviorally) a (psychological) preference for grapes. The preference for grapes in this economic sense rather than the desire for grapes in the FP sense is what neuroeconomics seeks to explain. It does not even aim to explain our ordinary concept of desire.

At the outset I said that this paper should not be read as an argument for eliminativism about FP states. The reason should now be clearer. I do not wish to argue here about whether agents undergo beliefs and desires, or whether beliefs and desires are real psychological states in a deep metaphysical sense. The relevant question as I see it is not whether those states exist but whether they figure into scientific explanations of choice behavior.<sup>9</sup> They do not, as least as far as neuroeconomics is concerned. And this is all the more reason to avoid the practice of forcibly mapping them onto the results of empirical science. Those empirical results are already mapped onto their relevant higher-level concepts, which are the concepts dictated by the raw empirical observations that serve as the target phenomena in need of explanation.

One possible strategy for proponents of FP to pursue in response to this argument is to make their case for the indispensability of FP at higher scientific levels. It seems to me that defenders of the explanatory relevance of FP states have generally tried to show that neuroscientific explanations preserve the conceptual framework of FP (e.g., Schroeder 2004). They might instead turn their attention to the relationship between FP and other higher-level theories such as EUT, rather than neuroscience. Neuroscience might have more to say about desires if desires in the FP sense figured into the higher level empirical theories from which neuroscience sets off.

But that strategy too is problematic. The problem is that the concept of utility in EUT more or less just is the attempt by scientists to capture the essence of our commonsense notion of desire. The economist Daniel Kahneman has recently been very explicit about this point:

---

<sup>9</sup> The view I am developing here shares some characteristics with what Bickle (2012) calls “little-e eliminativism”. His argument appeals not to neuroeconomics but rather to affective neuroscience.

As economists and decision theorists apply the term, it means “wantability”—and I have called it *decision utility*. Expected utility theory, for example, is entirely about the rules of rationality that should govern decision utilities; it has nothing at all to say about hedonic experiences. Of course, the two concepts of utility will coincide if people want what they will enjoy, and enjoy what they chose for themselves—and this assumption of coincidence is implicit in the general idea that economic agents are rational (Kahneman 2011, 377; original emphases).

Since the concept of utility is used to capture the quantifiable element of desire—the relative wantability of something; its likelihood of being chosen from among competing options—and it is this element rather than the concept of desire in the FP sense that serves these empirical purposes, it is hard to make sense of the demand that economists ought to do more with the concept of desire than they have. For that seems to amount to little more than the claim that the special sciences, including psychology and economics, have failed to take seriously our commonsense intuitions about the structure of the mind. If the sciences have failed in this regard, there are really only two possible explanations, neither of which will prove useful to proponents of FP. Either the special sciences have neglected those intuitions because those intuitions were flawed, somehow the products of confusion that require no explanation, or the special sciences have neglected those intuitions because they express facts about the world—raw data or observations—that the sciences simply cannot access.

Proponents of FP will obviously resist the first answer, since their goal is to defend the inevitability of FP. This leaves only the second answer as a real possibility. But what kind of raw data have economists and psychologists neglected due to inaccessibility? It cannot be observational data about preferences, since as we have just seen those are empirically accessible by way of behavior, and the concept of utility is built to organize observations of preferences. So it must then be a kind of introspective data that has been neglected. Psychology and economic theory, the proponents of FP claim, fail to explain our first-person experiences of our own mental states in the way we hoped or expected. Economic explanations of preferences ought to do more to accommodate what we know, through first-person subjective experience, to be true about ourselves, namely that we have desires (and, by extrapolation, so do others).

In this case the objection depends upon the assumption that first-

person introspective experience itself counts as the kind of raw data that simply cannot be ignored, whatever science says. As FP realists like Horgan and Graham put it, “our ordinary epistemic standards for folk psychological attributions are linked so closely to the truth or satisfaction conditions of such attributions that the truth of FP is beyond all serious doubt” (1990, 109). The thing to notice about this response is that, whether or not it works philosophically, it is characteristically anti-naturalistic in its insisting that there are some philosophical conclusions about human psychology that science could not ever overturn. Such a conclusion seems obviously incompatible with the pursuit of projects that involve attempt to show that neuroscience vindicates FP. If our first-person knowledge of desires is incontrovertible evidence of their existence then there is simply no point in looking to the sciences for support or vindication, as many philosophers are recently wont to do.

This brings me back to the broader implications for naturalism. Either the sciences sometimes guide the development of philosophical concepts or they do not. The latter position, which holds that the sciences do not guide conceptual development, privileges the traditional division of labor between philosophy and science. On this view we need not check disputes about philosophical theories against results in the sciences at all since those sciences fail to capture indisputable psychological facts. The former view, which holds that the sciences can in fact guide conceptual development in important ways, emphasizes the continuity between philosophy and science as a necessary feature of naturalism proper.

## CONCLUSION

The increasing empirical sophistication of philosophers, who now regularly appeal to scientific data for the purpose of constraining and adjudicating philosophical theorizing, is encouraging, especially to those of us who take philosophical and scientific inquiry to be continuous in their aims and methods. Still, I have argued, any such naturalistic approaches that leave unexamined the conceptual frameworks of traditional philosophical disputes are likely to fall short of genuine continuity. The case of decision-making and motivation is illustrative. The discontinuity between philosophical accounts of judgment and motivation, with their FP frameworks, and the developing empirical sciences of decision, such as neuroeconomics, shows not just

that we may have been too willing as philosophers to invent our psychology from scratch (Darwall, et al. 1992), but more importantly that we have so far paid too little attention to the ways in which the special sciences can themselves guide important conceptual developments.

In closing, I want to call attention directly to what many readers will regard as a loose end. This is the connection between *moral* philosophy and the special sciences. After all, I began by discussing the relationship between naturalism and philosophical *moral* psychology while much of the argument focused on the science of decision and motivation more generally. Earlier I argued that because moral cognition is a subset of cognition, there is no special reason to worry about moving from moral decision-making and motivation to decision-making and motivation more generally. Still, one might wonder why the connection is needed at all. What does moral philosophy really have to do with the preceding argument? I want to conclude with two brief remarks on this issue.

Firstly, the connection has a long and important history. A good deal of philosophical interest in the empirical psychology and neuroscience of motivation has come by way of traditional philosophical interest in morality and moral agency. Philosophical naturalism owes a significant debt to metaethics for developing the idea of continuity between philosophy and the sciences. The study of morality and moral agency was not so long ago squarely in the province of philosophy or theology (Doris and Stich 2007). Today moral philosophers play prominent roles in scientific laboratories and collaboration between philosophers and psychologists is commonplace. Perhaps channeling Hume, moral philosophers now spend a good deal of time drawing connections between human psychology and moral agency. As Hume saw, some of the most interesting questions about human psychology are at the same time questions about moral norms and one's relationship to others and the world around us. The question of whether it is contrary to reason for me to prefer the destruction of the whole world to the scratching of my finger has an unavoidably moral flavor. Because morality exists not in a philosophical vacuum but as a relationship between rational agents, it should not be altogether surprising (though it nevertheless shows remarkable prescience) that in the 1950s philosophers like Anscombe, as quoted in my epigraph, were already calling attention to the impossibility of genuine progress in ethics independently of progress in philosophical psychology.

Secondly, and relatedly, this paper's argument begins with a focus on moral philosophy precisely because it is in moral philosophy—in metaethics and philosophical moral psychology in particular—that we now find a rather useful philosophical framework for assessing philosophical presuppositions about human judgment, decision-making, and motivation against developing empirical claims. Disputes between Humeans and anti-Humeans, and motivational internalists and externalists are well suited to this paper's argument because they are prominent in the philosophical literature, a point of interest for scientifically minded philosophers, and yet open to empirical examination in the sense that their conceptual presuppositions can now be directly investigated by neuroscientists. Whether agents are necessarily motivated to choose in accordance with their beliefs turns out to be the kind of conceptually and empirically complicated question about which the special sciences have much to say. The questions about moral agency need not—as Anscombe, Darwall, and many other philosophers have argued, *should* not—be decided by unaided philosophical intuitions about human psychology. There are more resources from which to draw.

The central argument in this paper draws initially on moral philosophy not because it aims to show that scientific developments such as those in neuroeconomics speak to moral philosophers exclusively, but rather because moral philosophy has been, to its credit, a field rich with interest in the direct application of scientific data to traditional philosophical theorizing. As a result, it has done a great deal to advance our understanding of what philosophical naturalism amounts to in principle and in practice. The implication of my argument, then, is not that philosophical moral psychology is to be singled out for falling short of naturalistic standards but rather that in emphasizing the relationship between moral agency and rational agency it has given us a new vantage point on the old demarcation problem.

## REFERENCES

- Altham, James E. J. 1986. The legacy of emotivism. In *Fact, science and morality: essays on A. J. Ayer's Language, truth and logic*, eds. Graham Macdonald, and Crispin Wright. Basil Blackwell, 275-288.
- Anscombe, G. Elizabeth M. 1957. *Intention*. Basil Blackwell.
- Anscombe, G. Elizabeth M. 1958. Modern moral philosophy. *Philosophy*, 33 (124): 1-19.
- Bayer, Hannah M., and Paul W. Glimcher. 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47 (1): 129-141.



- Bickle, John. 2012. Finding the mechanisms of affect. In *Categorical versus dimensional models of affect: a seminar on the theories of Panksepp and Russell*, eds. Peter Zachar, and Ralph Ellis. Amsterdam: John Benjamins Publishing Co., 175-187.
- Darwall, Stephen, Alan Gibbard, and Peter Railton. 1992. Toward fin de siècle ethics: some trends. *The Philosophical Review*, 101 (1): 115-189.
- Dennett, Daniel. 1987. *The intentional stance*. Cambridge: MIT Press.
- Doris, John, and Stephen Stich. 2007. As a matter of fact: empirical perspectives on ethics. In *The Oxford handbook of contemporary philosophy*, eds. Frank Jackson, and Michael Smith. New York: Oxford University Press, 114-152.
- Gauker, Christopher. 2005. The belief-desire law. *Facta Philosophica*, 7 (2): 121-144.
- Glimcher, Paul W., Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack. 2009. Introduction: a brief history of neuroeconomics. In *Neuroeconomics: decision making and the brain*, eds. Paul Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack. London: Academic Press, 1-11.
- Glimcher, Paul W. 2009. Choice: towards a standard back-pocket model. In *Neuroeconomics: decision making and the brain*, eds. Paul Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack. London: Academic Press, 503-520.
- Greene, Joshua, and Jonathan Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6 (12): 517-523.
- Horgan, Terence, and George Graham. 1990. In defense of southern fundamentalism. *Philosophical Studies*, 62: 107-134.
- Jackson, Frank, and Philip Pettit. 1990. In defence of folk psychology. *Philosophical Studies*, 59 (1): 31-54.
- Kable, Joseph W., and Paul W. Glimcher. 2009. The neurobiology of decision: consensus and controversy. *Neuron*, 63 (6): 733-745.
- Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica*, 47 (2): 263-291.
- Kahneman, Daniel. 2011. *Thinking fast and slow*. New York: Farrar, Strauss and Giroux.
- Kennett, Jeanette, and Cordelia Fine. 2009. Will the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology. *Ethical Theory and Moral Practice*, 12 (1): 77-96.
- Moll, Jorge, Ricardo de Oliveira-Souza, Ivanei E. Bramati, and Jordan Grafman. 2002. Functional networks in emotional and nonmoral social judgments. *NeuroImage*, 16 (3A): 696-703.
- Padoa-Schioppa, Camillo, and John A. Assad. 2006. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441: 223-226.
- Padoa-Schioppa, Camillo, and John A. Assad. 2008. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, 11: 95-102.
- Prinz, Jesse. 2006. The emotional basis of moral judgments. *Philosophical Explorations*, 9 (1): 29-43.
- Quine, Willard Van Orman. 1969. *Ontological relativity and other essays*. New York: Columbia University Press.
- Roitman, Jamie D. and Michael N. Shadlen. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22 (21): 9475-9489.

- Roskies, Adina. 2003. Are ethical judgments intrinsically motivational? Lessons from 'acquired sociopathy'. *Philosophical Psychology*, 16 (1): 51-66.
- Roskies, Adina. 2006. A case study in neuroethics: the nature of moral judgment. In *Neuroethics: defining the issues in theory, practice, and policy*, ed. Judy Illes. Oxford: Oxford University Press, 17-32.
- Rustichini, Aldo. 2009. Neuroeconomics: formal models of decision making and cognitive neuroscience. In *Neuroeconomics: decision making and the brain*, eds. Paul Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack. London: Academic Press, 33-45.
- Samejima, Kazuyuki, Yasumasa Ueda, Kenji Doya, and Minoru Kimura. 2005. Representation of action-specific reward values in the striatum. *Science*, 310: 1337-1340.
- Sanfey, Alan, Mire Stallen, and Luke J. Chang. 2014. Norms and expectations in social decision making. *Trends in Cognitive Science*, 18 (4): 172-174.
- Schroeder, Timothy. 2004. *Three faces of desire*. Oxford University Press.
- Schroeder, Timothy. 2009. Desire. *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta.
- Schroeder, Timothy, Adina Roskies, and Shaun Nichols. 2010. Moral motivation. In *The moral psychology handbook*, ed. John Doris. Oxford University Press, 72-110.
- Schultz, Wolfram, Peter Dayan, and P. Read Montague. 1997. A neural substrate of prediction and reward. *Science*, 275: 1593-1599.
- Shadlen, Michael N., and William T. Newsome. 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, 86: 1916-36.
- Stich, Stephen. 1983. *From folk psychology to cognitive science: the case against belief*. Cambridge: MIT Press.

**Daniel Hartner** is assistant professor of philosophy at the Rose-Hulman Institute of Technology in Terre Haute, Indiana. His research focuses primarily on issues in neuroethics and the relationship between philosophy and neuroscience.

Contact e-mail: <hartner@rose-hulman.edu>