

The Case Against Economic Values in the Orbitofrontal Cortex (or Anywhere Else in the Brain)

Benjamin Y. Hayden¹ and Yael Niv²

¹ Department of Neuroscience, Center for Magnetic Resonance Research & Center for Neuroengineering, University of Minnesota

² Psychology Department & Princeton Neuroscience Institute, Princeton University

Much of traditional neuroeconomics proceeds from the hypothesis that value is reified in the brain, that is, that there are neurons or brain regions whose responses serve the discrete purpose of encoding value. This hypothesis is supported by the finding that the activity of many neurons covaries with subjective value as estimated in specific tasks, and has led to the idea that the primary function of the orbitofrontal cortex is to compute and signal economic value. Here we consider an alternative: That economic value, in the cardinal, common-currency sense, is not represented in the brain and used for choice by default. This idea is motivated by consideration of the economic concept of value, which places important epistemic constraints on our ability to identify its neural basis. It is also motivated by the behavioral economics literature, especially work on heuristics, which proposes value-free process models for much if not all of choice. Finally, it is buoyed by recent neural and behavioral findings regarding how animals and humans learn to choose between options. In light of our hypothesis, we critically reevaluate putative neural evidence for the representation of value and explore an alternative: direct learning of action policies. We delineate how this alternative can provide a robust account of behavior that concords with existing empirical data.

Keywords: neuroeconomics, decision making, value, orbitofrontal cortex, reinforcement learning

The past 20 years have seen a great deal of interest in understanding how our brains implement economic choices (Camerer et al., 2005; Glimcher & Fehr, 2013; Loewenstein et al., 2008; Padoa-Schioppa, 2011; Rangel et al., 2008; Rushworth et al., 2011). Much research in this field of neuroeconomics rests on the assumption that choices between options rely on an explicit valuation process (Kable & Glimcher, 2009; Levy & Glimcher, 2012; Montague & Berns, 2002; O'Doherty, 2014; Padoa-Schioppa, 2011). That is, that the brain first assigns value to each option and then compares those values to determine choice.

The concept of valuation—which stems from economic theory—is so ingrained that it may seem inevitable. How else would one literally compare apples to oranges? Indeed, the idea that value exists on a single cardinal scale, also called a “common currency,” has been extended to encompass not only goods, but also effort costs and time delays; everything an agent needs to bundle into evaluation of a mode of action that is aimed at procuring a specific outcome.

In parallel, the computational framework of reinforcement learning, which has been a cornerstone of neuroeconomics, also makes a scalar value signal central to its implementation (Niv, 2009; Sutton & Barto, 2018). Specifically, in reinforcement learning, the value of an option—a state or an action—is the expected sum of future rewards contingent upon that choice. As a sum of future rewards that may be of different types (and include costs as negative rewards), reinforcement-learning value is naturally calculated in some unitless common currency.

However, not all reinforcement learning algorithms rely on or even calculate values (Sutton & Barto, 2018), and reinforcement-learning values, as sums of future rewards, are not synonymous with economic values of specific goods. Likewise, many empirically supported process models of choice get by with no valuation (Gigerenzer & Gaissmaier, 2011; Miller et al., 2019; Vlaev et al., 2011). The fact that the brain *can* compute values to compare apples and oranges does not mean that it routinely does so, or that valuation is the primary process underlying choice. In this opinion paper, we argue that in many choice scenarios the brain may not be computing values at all, despite appearing to do so. We will demonstrate that the rationale for a value signal is weak, as are both behavioral and neural evidence supporting value computation. Finally, we propose an alternative—direct learning of action policies—and suggest there is

Yael Niv  <https://orcid.org/0000-0002-0259-8371>

This research was supported by National Institute on Drug Abuse Grants R01 DA038615 (B.Y.H.) and R01 DA042065 (Y.N.), and by grants from the John Templeton Foundation (to both Y.N. and B.Y.H.). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation or the National Institute on Drug Abuse.

The authors have no competing interests to declare.

We thank the Gordon Research Conference on Neurobiology of Cognition 2018, organized by Jennifer Groh and David Leopold, for providing the context in which this collaboration was seeded. We thank Sarah Heilbronner, Rei Akaishi, Becket Ebitz, and many commentators on Twitter for helpful discussions.

Correspondence concerning this article should be addressed to Yael Niv, Psychology Department & Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, United States. or to Benjamin Y. Hayden, Department of Neuroscience, University of Minnesota, Minneapolis, MN 55455, United States. Email: yael@princeton.edu or benhayden@gmail.com

scant direct evidence for value learning that cannot be explained by this and other alternative theories.

Our hypothesis is important as it suggests a different interpretation of previous data and requires that studies attempting to resolve mechanisms of value computation in the brain first establish that in the specific situation studied, valuation is indeed occurring. In particular, much work has associated the orbitofrontal cortex (OFC) in representing the expected economic value of different goods or options (Bartra et al., 2013; Levy & Glimcher, 2012)—a role that must be critically reconsidered if we agree that the brain may not necessarily be representing such values in many of the experiments so far used to test this hypothesis.

Why Argue Against Value?

Intuitively, our thesis is that while we may know that we prefer an orange to an apple (for one thing, oranges don't brown when exposed to oxygen), we may make this judgment without consulting an internal scalar (cardinal) or even a universal ordinal value signal. Consequently, we may be hard-pressed to express the precise value that we put on an orange. This is not merely an issue of conscious access to our internal value of oranges¹, but may be due to the fact that deciding on preferences can utilize many alternative mechanisms that don't require or rely on calculation of such a value.

Valuation is hard (Payne et al., 1992). It is also often unnecessary: when choosing between, say, an orange and a car, it is immediately clear that one is better than the other without calculating the precise value of either. If you are extremely thirsty, you might choose the orange, whereas if you need to go somewhere a car is the only relevant option. Arguably, many real-life choices are between options that are sufficiently different in the needs that they fulfill as to be more similar to this extreme example than they are to the choice between apples and oranges (Juechems & Summerfield, 2019). Valuation may only be necessary when choosing between two very similarly valued items. However, if the items are sufficiently similar in how they satisfy our needs, the brain may decide to choose randomly, according to some value-free heuristic, or according to past choices—and move on (Chater, 2018). Alternatively, the brain can try to calculate the exact values of the options to the necessary precision that arbitrates between them. Indeed, choosing between similar options often takes longer than choosing between very different options, even when making rather trivial choices that do not warrant the time and effort invested (Pirrone et al., 2018; Teodorescu et al., 2016). An example is the inordinate amount of time some of us may spend deciding what brand of tuna fish to buy, despite the fact that our time (as per our salary) is worth much more than the money we will save by correctly evaluating which brand provides the best “value for money.” This scenario also illustrates how inept we are at comparing the value of different kinds—here, time and money—which should have been straightforward if the *modus operandi* of the brain was to evaluate everything in terms of some common currency (see below).

So perhaps the brain *can*, when needed, calculate values. However, we argue that this is not the main means by which the brain makes decisions, and perhaps not the natural mode of decision making.

What Is Value?

Before moving on, we would like to delineate precisely how we are defining value for the sake of our argument. The term “value” has multiple uses (the problems this multiplicity raises are carefully laid out by O'Doherty, 2014). We consider “value” a hypothesized scalar variable that reflects the worth of a specific item or outcome.² Because it is scalar, it is necessarily abstract. It can refer to any good, and makes use of a common currency code that is comparable across goods of different types (e.g., food, water, recreation time). Value, by this definition, is cardinal, not ordinal, meaning it can be defined for an option *per se*, rather than solely relative to other options. That is, it reifies the idea of “utils”—quantifiable units of value, often used in a jocular manner in Economics classes.

The idea of a common currency is key as it implies that all relative calculations have already happened—the value of an option is in some denomination that objectively defines it as compared to other such values. One might argue that this is too narrow a definition, but we believe this is what neuroeconomists have in mind when talking about representation of common currency values in the brain. For example, we know of no neuroeconomic model that imagines a neural implementation of an ordinal value scale. This having been said, even an ordinal scale should not change based on the comparison set, and should satisfy transitivity—which many of the neural signals attributed to value do not, as we detail below.

Our definition of value, for the purpose of this paper, is different from other types of value that are discussed in reinforcement learning. Specifically, here we are discussing value as *the reward worth of a single item/event*, not the expected sum of all future rewards (R in reinforcement learning models, not V). One might argue that in order to compute such an expected sum V , the subjective worth of all individual rewards must be translated to a common currency that can be added. In this sense, reinforcement learning models do presuppose a common currency reward value. However, they don't necessarily commit to economic properties of such values, such as transitivity and consistency (see also Juechems & Summerfield, 2019), and we discuss below a class of reinforcement learning algorithms that can make do with evaluations that are only relative to the current options, and not applicable in other situations.

One important feature of our definition of value is that in our view, although value is inferred from choice, it is not strictly identical to choice, nor necessarily implied by choice. For instance, if we observe a consistent preference for A over B, A is assumed to have a higher value than B. But choice of A over B isn't sufficient to infer the existence of a self-consistent value function: a decision-maker may adhere to a heuristic policy that results in stable preference for $[A > B]$, $[B > C]$, and $[C > A]$ (Lichtenstein & Slovic, 2006). Moreover, choice can be altered without manipulating the reward value of an item (Schonberg & Katz, 2020), which suggests that value, as inferred from choice, is not untarnished by processes that are not economic in nature. Thus, value can be inferred from behavior given certain assumptions, but preferences do not always lead to a value function.

¹ We note here that, in neuroeconomics, value-based choices are generally thought to be part of explicit, aware, goal-directed, or model-based decision making that relies on frontal-cortex areas, in particular, the OFC.

² We consider subjective worth as revealed by preference, following neuroeconomic theory, but note the limiting focus on only the “wanting” not the “liking” side of value (Berridge, 1996).

The Common-Currency Hypothesis

The notion that economic decision-makers make use of internal value functions to compare options is often dated to Daniel Bernoulli's proposal in 1738 of a logarithmic utility curve to explain preferences in the St. Petersburg Paradox (Martin, 2011). Here, a decision-maker is offered a chance to play a game in which they win $\$2^{(n-1)}$ with n being the first time, in a series of coin tosses, that a coin falls on "heads." The paradox is that although the expected monetary value of this game (that is, the sum of expected wins multiplied by their probability $\sum_{n=1}^{\infty} 2^{(n-1)} \cdot 0.5^n = \sum_{n=1}^{\infty} 0.5$) is infinity, people are not willing to pay even \$10 to play this game (Hayden & Platt, 2009). The explanation that Bernoulli proposed proved foundational within microeconomic theory. He argued that decision-makers don't base calculations on the nominal, objective cash value of the potential gain, but rather on its subjective value. If the subjective value grows more slowly than the objective value (the idea of "diminishing marginal utility"), an optimizing decision-maker will appear risk-averse, for instance, when evaluating the St. Petersburg gamble.

Utility, or subjective value, has been central to many if not most microeconomic models. These include the axiomatic approaches of Pareto, Von Neumann and Morgenstern, and Samuelson. It is also central to behavioral theories such as prospect theory and decision field theory (Busemeyer & Townsend, 1993; Kahneman & Tversky, 1979). Ironically, however, explaining choices in the St. Petersburg paradox using such a subjective value function requires utility for money that diminishes so rapidly that it does not generalize to choices in other contexts. Instead, heuristic accounts—accounts that don't rely on ideas about value maximization—provide better quantitative matches for St. Petersburg choices (Hayden & Platt, 2009).

Despite its central importance in economics, economists are typically agnostic about whether the concept of value is just a convenient description, or whether it is instantiated in the brain. As Friedman argued, decision-makers behave *as if* we compute and compare values, but we cannot conclude that we actually do so (Friedman, 1953). In that work, he famously compared economic agents to a trained billiards player who makes excellent shots as if having a sophisticated grasp of Euclidean geometry, although in fact such a theoretical understanding is not necessary for good billiards skill.

Economists typically stop at the point of saying that economic models are "as if" models, with some arguing that the question of the underlying reality of economic variables is outside the domain of economics (Gul & Pesendorfer, 2008; Harrison, 2008). In contrast, neuroeconomics has generally taken as a default assumption that these "as if" theories are reified in the brain (Kable & Glimcher, 2009; Levy & Glimcher, 2012; Montague & Berns, 2002; O'Doherty, 2014; Padoa-Schioppa, 2011; Rich & Wallis, 2016).

Modern tools such as neuroimaging and single unit physiology provide the opportunity to assess the implementation of decision making directly. Indeed, neuroscientists have had little trouble identifying correlates of value in several brain areas (Knutson et al., 2001; Levy & Glimcher, 2012; Plassmann et al., 2007; Wallis, 2007), leading to the suggestion that the OFC is a nexus for representing the economic value of goods in the brain (see, e.g., Rushworth et al., 2011; Wallis, 2007), alongside other brain areas that are important for computing and representing value such as the ventromedial prefrontal cortex (Bartra et al., 2013) and ventral striatum (Haber & Knutson, 2010). However, it is not clear that the signals identified in these studies actually represent value as

proposed—on a common-currency scale, for comparing options and making choices. Moreover, so called "value signals" only *correlate* with value, so may be signaling other quantities, such as attention, action plans, vigor, or preference, which also correlate with value (Maunsell, 2004; O'Doherty, 2014; Wallis & Rich, 2011). Hence it is important to carefully consider alternative interpretations for these findings, as we do further below, after discussing some practical constraints and philosophical conundrums.

Alternatives to Valuation in Decision Making

While common-currency value provides a convenient and general mechanism for making choices, there are many possible alternatives (Gigerenzer & Gaissmaier, 2011; Kahneman et al., 1982; Lichtenstein & Slovic, 2006; Vlaev et al., 2011), some of which are quite general and robust. Consider for example the "priority heuristic" (Brandstätter et al., 2006). This *de minimis* heuristic approach proposes that decision-makers first identify a dimension along which options vary and then compare options along that dimension. If that results in a choice, they stop; otherwise, they move on to the next dimension. This heuristic can explain many phenomena, including Allais preferences (Allais, 1953), the reflection effect (Fishburn & Kochenberger, 1979), the certainty effect (Kahneman & Tversky, 1979), the fourfold pattern in risky choice that motivates prospect theory (Tversky & Fox, 1995), and several intransitivities—all without ever requiring computing of value (Brandstätter et al., 2006). And the priority heuristic is one of a large number of heuristics that do a remarkable job at describing behavior (Gigerenzer & Gaissmaier, 2011). These heuristics are generally motivated by psychological observations, and thus are consistent with known data from the psychology—if not the neuroscience—of choice. In particular, they reflect the assumption that calculating value is difficult, that humans typically use shortcuts whenever possible, and that heuristics are a good shortcut (Gigerenzer & Gaissmaier, 2011; Lieder & Griffiths, 2020; Payne et al., 1992). Moreover, these heuristics are not limited to humans, but apply to other species, including monkeys (Heilbronner & Hayden, 2016; Marsh, 2002; Santos & Rosati, 2015; Shafir et al., 2002).

Importantly for our argument, the success of heuristic approaches demonstrates that value calculations are not *a priori* essential for neuroeconomic theories of choice. In particular, heuristic theories can solve many problems for which value is proposed to be needed (Kahneman et al., 1982; Lichtenstein & Slovic, 2006; Piantadosi & Hayden, 2015; Stevens, 2016; Tversky, 1969; Vlaev et al., 2011). Heuristics readily allow for comparison of multi-dimensional goods and bundles, and for choice across dissimilar goods. While heuristics are not perfect, and lead to many choice anomalies, choice is, empirically, full of anomalies. Moreover, several non-heuristic process models also eschew value computation steps, including decision by sampling, query theory, and fuzzy trace theory; these all account for a wide range of choice behavior as well (Reyna, 2008; Stewart & Simpson, 2008; Stewart et al., 2006; Weber et al., 2007).

Practical Issues in the Neuroscience of Value Representation

Despite the above, it is often considered axiomatic that an internal, neural, value scale must exist in the brain. The job of neuroscientists is then to find this signal—to pinpoint brain activity

that correlates with value. In this section, we challenge this viewpoint by considering some basic issues that come up in the neuroscience of value.

The first problem is that it is impossible to precisely measure value. We can measure preferences between options and use the data to *infer* option values, but elicited preferences are noisy measurements that also reflect factors other than value. For example, depending on how preferences are elicited, they may also reflect the tendency to press the same button repeatedly rather than change actions, the tendency to switch between options due to a prior belief about depleting resources, or the amount of attention or looking time for each of the options (Armel et al., 2008; Schonberg & Katz, 2020; Shimojo et al., 2003; Sugrue et al., 2005). Finally, value may shift from trial to trial, even depending on recent outcomes, meaning that methods that average across trials produce misleading value estimates (Sugrue et al., 2005). One could, in theory, incorporate these factors into the inferred value of an option. For example, there may be inherent value in choosing the same thing twice in a row. However, it becomes unclear what the definition of the “option” is that is being evaluated: if the value of an apple after eating (or choosing, or even just viewing) an orange is different from the value of an apple without that proximal experience, can we determine the value of goods at all? And if we can change the value (read: preference) for an option just by directing more attention to it (Salomon et al., 2018; Schonberg et al., 2014), or inducing choice of it (Izuma et al., 2010; Sharot et al., 2012; Voigt et al., 2017), is preference really measuring the subjective economic worth of a good? The risk is circularity—if any choice behavior can be explained by supposing value that depends on local history, then the concept of value adds no additional explanatory power beyond that of recent events and choices.

Different ways to measure value raise a second challenge of inconsistent measures. Indeed, if we had a single neural value function we called on, values elicited by different measures would match. Unfortunately, they do not (Lichtenstein & Slovic, 2006). To explain this, in their seminal work, Sarah Lichtenstein and Paul Slovic proposed that preferences do not arise from any internal value function, but instead are constructed at the time of elicitation (see also Ariely et al., 2003; Payne et al., 1992; Tversky & Shafir, 1992). That is, in the view of these and like-minded scholars, value doesn't sit in the brain waiting to be used; rather, preference is a complex and active process that takes place at the time the decision is made. Critically, in this theory we compute choices in a largely ad-hoc manner based on the available options, without an intermediary common-currency valuation stage. As such, there is no guarantee of consistency or reliability; any consistency or reliability observed may be explained as a result of strong attractor states in the way the system determines the choice, and deviations from consistency are evidence for the specific nature of the algorithm and its idiosyncrasies. This hypothesis, while denouncing value as a latent construct in the brain, nevertheless invites neuroscientific research to understand the active processes involved in choice, and how these relate to (relative) evaluation, if not to economic values.

This having been said, experimenters *can* roughly estimate value, even if not measure it precisely. For example, in a given experiment they can examine choices and determine with confidence that a monkey (behaves as if it) places more value on a gamble as compared to a safe option with a matched expected value. One could then use this fact to try to identify a neural correlate of value.

However, for this neural endeavor to be valid, it is important to identify all confounding variables and regress them out. O'Doherty's (2014) review delineates the practical difficulty of doing so, using the overall rubric of “visceral, autonomic, and skeletomotor” activity. In practice, confounding variables include both stimulus and outcome identity, information about the state or structure of the world, the surpriseness, informativeness, and informational value of stimuli, details of the action associated with selecting of consuming the reward, including its likelihood and vigor, and the attention and arousal engendered by the stimulus (e.g., Blanchard et al., 2015; Botvinik-Nezer et al., 2020; Niv et al., 2007; O'Doherty, 2007; Roesch & Olson, 2004; Roesch et al., 2006; Wilson et al., 2014; Yoo et al., 2018). Indeed, in studies that separately assess encoding of outcome identity versus outcome value, activity in brain areas that are often considered to be emblematic of economic value (in particular, the OFC) turns out to correlate with outcome identity instead (Klein-Flügge et al., 2013).

A final insurmountable problem is that it may be impossible, even in theory, to obtain a brain measure of value that is independent of behavior. Suppose for example, that we identify a particular class of neurons whose firing rates are perfectly correlated with value down to our ability to measure it through preference. Supporting this idea, we observe that any procedure that modifies value (as inferred from behavior) changes the firing rate of these neurons in a manner consistent with our predictions. We may tentatively hypothesize that these neurons are (or are among) the value neurons of the brain. However, as shown by Schonberg and colleagues, preference can be changed irrespective of changing the economic worth of goods (Schonberg & Katz, 2020). Therefore, to differentiate value neurons from preference neurons, we would need to show that these neurons *do not* strictly follow expressed preference when the value function diverges from it. This is, of course, not possible if the value function never *measurably* diverges from preference. And if we assume that values *can* diverge from preference, it is not clear how to define values to start with. We call this “the neuroeconomic relativity problem” because, like Einstein's relativity problem, it reflects the fact that there is no external reference frame to which one can calibrate value inferences.

Reconsidering the Motivation for Common Currency

It may be worth asking, then, what does having a single common currency buy you? Why would the brain invest in such an organization? One advantage of a common-currency scale is that it simplifies comparing options that differ along multiple dimensions. For example, when hunting for an apartment, the options may differ along dimensions of price, area, neighborhood, and amenities. The logic is that these dimensions must be first combined into a single scalar per apartment so that the scalars can be compared. However, this is not the only way to solve the problem, and importantly, may not be the way humans make their decisions. For example, the apartment shopper may choose a single dimension and pick the winner along that dimension, as discussed above, or may compare separately on each dimension and choose the apartment that wins on most counts (Tversky, 1972). Laboratory studies where humans can choose what attributes to view indeed suggest that people don't uncover all the attributes of one option (to calculate its value) and then continue to the next option, but rather prefer to view information for all options attribute by attribute (Fellows, 2006; Hunt et al., 2014). Indeed, as mentioned, much empirical work indicates that human decision-makers broadly favor heuristic approaches that

eschew a value stage in a large number of contexts (Brandstätter et al., 2006; Gigerenzer & Gaissmaier, 2011; Kahneman et al., 1982; Lieder & Griffiths, 2020).

Notably, systems that make use of heuristics may generate internal variables that are conceptually distinct from value but that correlate with value, thus leading to an interpretational confound. Consider, for example, a relatively well-understood implementation of choice in a (non-brain) distributed system: the selection of hive sites in bees (*Apis mellifera*, Seeley, 2010; Seeley & Buhrman, 1999; Seeley et al., 2006). Bee swarms select a hive site by sending out scouts to investigate potential sites. Each site differs along roughly 20 dimensions of varying but measurable importance (size, safety from predators, exposure to sunlight, wind, etc.). Each scout bee that encounters the hive site performs an extremely poor estimation of the value of the site—it typically will only sample three-to-four of the dimensions and even then, estimate them poorly. The bee then returns and if the estimated site quality is sufficiently good, indicates an assessment of its quality to the other bees in the swarm. The quality it signals will be correlated with the overall value of the hive site, but only weakly, and, critically, will only integrate a subset of relevant value components. The overall comparison between options is indirect—the options race to attract adherents; the majority of the decision is made by a positive feedback process. This example is particularly relevant because scholars of perceptual decision making may recognize this process as strongly related to ideas about how the brain decides between options by racing to bounds, and there is evidence that deliberative processes in the brain follow the same principles as choice processes in bee swarms (Eisenreich et al., 2017; Franks et al., 2003; Mitchell, 2009; Pais et al., 2013; Passino et al., 2008; Pirrone et al., 2018).

Evaluating the Neural Evidence for the Common-Currency Hypothesis

Evidence for the common-currency hypothesis comes from the observation that firing rates of single neurons or hemodynamic responses of voxels correlate with values of offers and outcomes (Kennerley et al., 2009, 2011; Levy & Glimcher, 2012; Padoa-Schioppa, 2011; Rangel et al., 2008). These responses depend on multiple elements of offers (e.g., the expected reward, as well as the associated response costs), and are modulated by factors that affect subjective value, such as context or level of satiety (see Azab & Hayden, 2018; Conen & Padoa-Schioppa, 2019; O'Neill & Schultz, 2010; Rudebeck et al., 2017 for only a few examples). Often, variations in these firing rates predict variations in choice (Conen & Padoa-Schioppa, 2015; Strait et al., 2014; Sugrue et al., 2004). Such patterns have been taken as evidence that the neural signal encodes value, in particular implicating the OFC (Bartra et al., 2013; Levy & Glimcher, 2012; O'Reilly, 2020).

However, such patterns of neural responding do not definitively demonstrate a common-currency value signal. For example, in the original finding of value-encoding neurons in the OFC (Tremblay & Schultz, 1999), neurons responded more strongly to a high-valued option than to a low-valued one. However, the principle of common currency only held within a given choice pair—responses for Option B were low when this option was paired with a better Option A, but were high when this same option was paired with a worse Option C. Other studies found value-correlated responses to be more stable across contexts (Padoa-Schioppa & Assad, 2008), but most findings

support an encoding that depends on the alternatives (e.g., Kobayashi et al., 2010; Padoa-Schioppa, 2009; Zimmermann et al., 2018). One might interpret this finding as reflecting simple range adaptation of neurons that have a finite firing rate; however, a more parsimonious explanation is that the firing pattern is consistent with a relative preference code rather than an abstract value code. This interpretation places the putative neural representation of value one stage later than we would expect from a common-currency code—immediately after comparison (because a relative valuation is itself a comparison) rather than as an input to comparison, raising the question of where the input came from. An interpretation of this code as something other than economic value would obviate this worry. Importantly, with a code that depends on alternatives, there is no real sense in which we can read out the “true subjective value” of an option. We can only know if an option’s value is higher than another value—similar to the information provided by preferences and choice behavior.

Another corollary of such a relative code is that it is unclear to what extent we can read out a meaningful value signal when only one option is available. Indeed, presenting a subject with one option at a time should, theoretically, provide the best ability to read out option values from neural activity in areas associated with value, such as OFC, ventromedial prefrontal cortex, and dorsal and subgenual anterior cingulate cortices. Doing so reveals that while neural activity in these regions does correlate with the value of the first option, neural responses to the second (alternative) option (presented on its own) correlate with the value *difference*, that is, they reflect the result of value comparison rather than valuation per se (Azab & Hayden, 2018; Hunt et al., 2018; Strait et al., 2014). Moreover, even responses to the first offer do not simply encode its value, but also contain information about the likelihood that option will be chosen (presumably relative to the expected value of the second offer, Azab & Hayden, 2017). These results suggest that even when evaluation is experimentally segregated from comparison, pure value encodings may not exist. In fact, the idea that prefrontal neurons are selective or responsive to a single experimenter-defined variable has been increasingly falling out of favor (Fusi et al., 2016; Raposo et al., 2014; Rigotti et al., 2013), with “mixed selectivity” appearing to be a core operating principle of prefrontal cortex, including in ostensible value regions (Blanchard et al., 2018; Hayden & Platt, 2010; Kimmel et al., 2020; Yoo & Hayden, 2018).

A third issue is that although the activity of some neurons in the OFC is correlated with value (but not necessarily linearly related to value), activity of other neurons in this same area is anticorrelated with value, with the majority of neurons showing no relationship at all with value. This raises the possibility that the neurons are encoding something other than value, for instance, a distributed representation of the *identity* of each of the three options, A, B, and C above, or the identity of the stimulus representing the offer. With the small number of options evaluated in most experiments, such a distributed code of outcome or stimulus identity (but not value) can easily result in some neurons randomly firing most strongly for the higher-valued of the three options and least for the worse option, while in other neurons the relationship would be in the opposite direction. This would also explain why many neurons show a non-monotonic relationship between value and their firing.

Recently, using ensemble recordings in the OFC that allow analysis on the level of a single trial, Wallis and colleagues have attempted a more direct test of the hypothesis that OFC encodes

value (Rich & Wallis, 2016). In their task, monkeys were offered two options (denoted by images corresponding to the options) and asked to choose between them. On some trials, only one option was offered. Classifiers were trained to classify options corresponding to each of four offer-value levels (0.05, 0.10, 0.18, and 0.30 ml juice, or, in separate blocks, four levels of second-order reinforcer), aggregating over the two images corresponding to each value level. Using single-unit activity and local-field-potential recordings in the OFC, the authors could classify the offered value above chance, suggesting that OFC neurons encoded information about value. However, here too, offer values may have been encoded as different (outcome stimulus) identities, not different (cardinal) values in a common currency. One stringent test of the value-coding hypothesis would be to show that a combination of two offers of value level 0.05 ml resulted in similar activation pattern to a single offer of value level 0.10 ml. Ensemble recordings that give ample data in a single trial allow testing these hypotheses with novel combinations that have not been trained to predict the same identity of reward through multiple presentations, therefore testing the assumptions of the additivity of common currency directly. This critical test has not been conducted, to our knowledge.

As mentioned, many factors that are conceptually distinct from value can influence choice—and are therefore closely correlated with value (Maunsell, 2004). Reward value drives attention, promotes both short- and long-term learning, primes behavioral adjustments, updates internal models, activates circuitry that detect both positive and negative surprises, and elicits mental computations of cost–benefit tradeoffs as well as comparisons with what could have been chosen. All of these are different from scalar, common-currency value, but are known to drive neural activity in the regions usually associated with economic value. As such, they confound that interpretation. This problem is a long-standing and notorious one in neuroeconomics (Maunsell, 2004; O’Doherty, 2014; Roesch & Olson, 2003).

To overcome some of these potential confounds, a strong tradition in neuroeconomics research is to control for extraneous factors such as salience and response cost (although we note that many factors—for example, covert attention—cannot be controlled for). As a practical issue, it is difficult to control for alternative interpretations without making the task so convoluted that it becomes unnatural for the animal, in the ethological sense (O’Doherty, 2014). As a result, it is not clear if findings of value computation (if we believe them to be so) in these tasks would naturally translate to how the brain computes choice in naturalistic situations. But the deeper issue remains that value as inferred in most experiments is definitionally a summary of aggregated choice behavior, and therefore, any variable that influences choice behavior will necessarily be correlated with value. It is for this reason that we suggest that a stronger demonstration of scalar value coding in the brain should show mathematical properties such as the additivity of value, or separation from preferences as these are changed without changing value, as suggested above. If the response of neurons to the novel sum of two stimuli each promising 0.1 ml of juice is similar to the response of those same neurons to a different stimulus associated with 0.2 ml of juice, and especially if this response did not change when preference for an option was induced by means such as the mere exposure effect (Schonberg & Katz, 2020), it would be harder to argue that this is due to a shared motor plan, or attentional capture. We note that research to date has shown that the activity of brain

areas associated with value, in particular the OFC, *does* change when preferences are modified through methods that should, in principle, not change economic value (Botvinik-Nezer et al., 2020).

An Alternative View: Direct Learning of Policies

Although reinforcement-learning models have contributed to the assumption that the brain computes values, many reinforcement-learning algorithms do not learn or estimate values for different actions. Instead, they directly learn action policies. In fact, in the reinforcement-learning literature, the goal of an agent is to obtain as much reward as possible by executing optimal actions—calculating values is only one means to achieve that end.

In explaining reinforcement-learning methods, Dayan and Abbott’s (2001) textbook begins with the “direct actor”—an actor that learns actions without computing their values. The Actor-Critic model, a prominent algorithm for reinforcement learning that has been linked to the brain (Barto, 1995; Joel et al., 2002; Maia, 2010; O’Doherty et al., 2004; Takahashi et al., 2008), does exactly that. In this algorithm, a Critic module learns values of states in terms of expected future rewards (here, the state includes all available actions, averaging over choices and explicitly *not* computing the value of each possible choice), and uses these to compute reward prediction errors. These prediction errors are used to learn an action policy in the Actor module: the probability of actions that are followed by positive prediction errors is increased, and the probability of actions that are followed by negative prediction errors is decreased. Under some reasonable conditions, this model learns correct reward-maximizing policies (Sutton & Barto, 2018). However, the quantities learned by the Actor—tendencies to perform one action over the other—cannot be read out as action values. In particular, due to the way the algorithm learns, ties are broken between equally good actions such that eventually agents learn deterministic policies. To be clear, if four actions were to lead to 1, 2, 3, and 3 drops of juice, respectively, the model may learn to always choose the third option, or to always choose the fourth (both optimal policies). At the end of learning, action weights in the Actor may be 0, 0, 1, 0 respectively, losing all information about value, or relative value. Even before convergence to a deterministic policy, weights may be 0, 0.1, 0.95, 0.2 respectively (or any other combination—weights here do not have to sum up to 1, and these are just illustrative numbers). Importantly, if the basal ganglia indeed implement an Actor-Critic learning algorithm in the brain, there is no sense in which we can glean action values from the Actor or the Critic.

The Actor-Critic model is only one of a class of reinforcement-learning algorithms that learn policies directly, that is, without calculating option values (e.g., Sutton et al., 2000; Williams, 1992). In their general form, these algorithms maintain an action policy, use experience to evaluate a gradient direction for this policy (that is, what change in policy would increase the overall obtained reward), and change the policy in that direction. Another notable model, developed to explain behavioral patterns in choices, is the Experience-Weighted Attraction model of Camerer and Ho (1999), which interpolates between value calculation and direct policy learning.

Recent findings suggest that both behavior and learning signals measured in humans during decision making are more in line with a policy-learning algorithm than with value estimation. For instance,

Li and Daw (2011) had participants choose between two options that gave reward with different probabilities. After each choice, both the outcome of the chosen option and the counterfactual outcome of the unchosen option were displayed. Behavior, as well as neural signals corresponding to prediction errors in the basal ganglia, suggested that subjects were updating both options in opposite directions, learning relative choice propensities (a policy) rather than tracking the expected value of each option. In another task in which humans chose between pairs of options and were able to view both the outcome of their choice and the counterfactual outcome of the forgone option, Palminteri et al. (2015) had subjects learn which of two probabilistically rewarding options is better, and which of two probabilistically punishing options was better. At test, subjects were asked to choose between pairs from both the rewarding and the punishing contexts. Surprisingly, when choosing between the less rewarding option and the less punishing option, subjects tended to choose the less punishing option. This is consistent with policy learning, as that option had been the favored option in the punishment context, whereas the less rewarding option had been the disfavored one in the reward context. However, the value of a sometimes-rewarding option is clearly higher than that of a sometimes-punishing option, hence value learning cannot explain this fundamentally suboptimal choice pattern.

Direct learning of policies is consistent with basic tenets of decision making, such as the fact that choices are stochastic even when no exploration is necessary or warranted (e.g., choosing between two gambles that are fully described; Khaw, Li, et al., 2017). Because policies are relative quantities (preference for one option implies unfavorability of another, as the probabilities of all choices have to sum to one), they also explain common violations of the independence axiom such as the effects of third-option “decoys” on choice (Soltani et al., 2012), and temporal and spatial contextual influences on choice (Khaw, Glimcher, et al., 2017), although these phenomena can also be explained by valuation models that involve range normalization.

Conclusion

The field of neuroeconomics started with the putative identification of pure value signals (Platt & Glimcher, 1999). The meaning of these signals was disputed early on (Maunsell, 2004; Roesch et al., 2006), but deep questions were set aside as researchers continued to identify brain areas with parts of the computational processes of choice, and in particular, identified the OFC with the seat of economic value. However, while those debates have subsided, the problems they raised have not been resolved. Progress on these issues will require additional data, but we stress that not every experiment that involves choices also implies valuation, and one must be careful in interpreting data not only because of potential confounds, but also because we must be wary of treating a hypothesis—that the brain computes value—as an axiom. We also suggest the need for additional philosophical work to define value in a way that is—at least in principle—dissociable from other factors that promote choices (Juechems & Summerfield, 2019). In other words, we argue for a return to the productive debates of the early days of the field, 20 years ago (for neuroeconomics), and earlier (for its psychological underpinnings). Bolstered by an additional 20 years of new data, such debates would surely benefit the field of neuroeconomics moving forward. They would also potentially help

reconcile conflicting views on what the OFC might or might not be doing in decision making (Stalnaker et al., 2015).

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica*, 21, 503–546. <https://doi.org/10.2307/1907921>
- Ariely, D., Loewenstein, G., & Prelec, D. (2003). “Coherent arbitrariness”: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1), 73–106. <https://doi.org/10.1162/00335530360535153>
- Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, 3(5), 396–403.
- Azab, H., & Hayden, B. Y. (2017). Correlates of decisional dynamics in the dorsal anterior cingulate cortex. *PLOS Biology*, 15(11), Article e2003091. <https://doi.org/10.1371/journal.pbio.2003091>
- Azab, H., & Hayden, B. Y. (2018). Correlates of economic decisions in the dorsal and subgenual anterior cingulate cortices. *European Journal of Neuroscience*, 47(8), 979–993. <https://doi.org/10.1111/ejn.13865>
- Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215–232). MIT Press.
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412–427. <https://doi.org/10.1016/j.neuroimage.2013.02.063>
- Berridge, K. C. (1996). Food reward: Brain substrates of wanting and liking. *Neuroscience and Biobehavioral Reviews*, 20(1), 1–25. [https://doi.org/10.1016/0149-7634\(95\)00033-B](https://doi.org/10.1016/0149-7634(95)00033-B)
- Blanchard, T. C., Piantadosi, S. T., & Hayden, B. Y. (2018). Robust mixture modeling reveals category-free selectivity in reward region neuronal ensembles. *Journal of Neurophysiology*, 119(4), 1305–1318. <https://doi.org/10.1152/jn.00808.2017>
- Blanchard, T. C., Strait, C. E., & Hayden, B. Y. (2015). Ramping ensemble activity in dorsal anterior cingulate neurons during persistent commitment to a decision. *Journal of Neurophysiology*, 114(4), 2439–2449. <https://doi.org/10.1152/jn.00711.2015>
- Botvinnik-Nezer, R., Salomon, T., & Schonberg, T. (2020). Enhanced bottom-up and reduced top-down fMRI activity is related to long-lasting nonreinforced behavioral change. *Cerebral Cortex*, 30(3), 858–874. <https://doi.org/10.1093/cercor/bhz132>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113(2), 409–432. <https://doi.org/10.1037/0033-295X.113.2.409>
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459. <https://doi.org/10.1037/0033-295X.100.3.432>
- Camerer, C., & Ho, T.-H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4), 827–874. <https://doi.org/10.1111/1468-0262.00054>
- Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, 43(1), 9–64. <https://doi.org/10.1257/0022051053737843>
- Chater, N. (2018). *The mind is flat: The remarkable shallowness of the improvising brain*. Yale University Press.
- Conen, K. E., & Padoa-Schioppa, C. (2015). Neuronal variability in orbitofrontal cortex during economic decisions. *Journal of Neurophysiology*, 114(3), 1367–1381. <https://doi.org/10.1152/jn.00231.2015>
- Conen, K. E., & Padoa-Schioppa, C. (2019). Partial adaptation to the value range in the macaque orbitofrontal cortex. *The Journal of Neuroscience*, 39(18), 3498–3513. <https://doi.org/10.1523/JNEUROSCI.2279-18.2019>

- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems* [Computational neuroscience series]. MIT Press.
- Eisenreich, B. R., Akaishi, R., & Hayden, B. Y. (2017). Control without controllers: Toward a distributed neuroscience of executive control. *Journal of Cognitive Neuroscience*, 29(10), 1684–1698. https://doi.org/10.1162/jocn_a_01139
- Fellows, L. K. (2006). Deciding how to decide: Ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain: A Journal of Neurology*, 129(4), 944–952. <https://doi.org/10.1093/brain/awl017>
- Fishburn, P. C., & Kochenberger, G. A. (1979). Two-piece von Neumann–Morgenstern utility functions. *Decision Sciences*, 10(4), 503–518. <https://doi.org/10.1111/j.1540-5915.1979.tb00043.x>
- Franks, N. R., Dornhaus, A., Fitzsimmons, J. P., & Stevens, M. (2003). Speed versus accuracy in collective decision making. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 270(1532), 2457–2463. <https://doi.org/10.1098/rspb.2003.2527>
- Friedman, M. (1953). The methodology of positive economics. *Essays in Positive Economics*, 3(3), 145–178.
- Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66–74. <https://doi.org/10.1016/j.conb.2016.01.010>
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482. <https://doi.org/10.1146/annurev-psyich-120709-145346>
- Glimcher, P. W., & Fehr, E. (Eds.). (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.
- Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. *The Foundations of Positive and Normative Economics: A Handbook*, 1, 3–42.
- Haber, S. N., & Knutson, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35(1), 4–26. <https://doi.org/10.1038/npp.2009.129>
- Harrison, G. W. (2008). Neuroeconomics: A critical reconsideration. *Economics and Philosophy*, 24(3), 303–344. <https://doi.org/10.1017/S0266267108002009>
- Hayden, B. Y., & Platt, M. L. (2009). The mean, the median, and the St. Petersburg paradox. *Judgment and Decision Making*, 4(4), 256.
- Hayden, B. Y., & Platt, M. L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *The Journal of Neuroscience*, 30(9), 3339–3346. <https://doi.org/10.1523/JNEUROSCI.4874-09.2010>
- Heilbronner, S. R., & Hayden, B. Y. (2016). The description–experience gap in risky choice in nonhuman primates. *Psychonomic Bulletin & Review*, 23(2), 593–600. <https://doi.org/10.3758/s13423-015-0924-2>
- Hunt, L. T., Dolan, R. J., & Behrens, T. E. (2014). Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience*, 17(11), 1613–1622. <https://doi.org/10.1038/nn.3836>
- Hunt, L. T., Malalasekera, W. N., de Berker, A. O., Miranda, B., Farmer, S. F., Behrens, T. E., & Kennerley, S. W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. *Nature Neuroscience*, 21(10), 1471–1481. <https://doi.org/10.1038/s41593-018-0239-5>
- Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., & Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51), 22014–22019. <https://doi.org/10.1073/pnas.1011879108>
- Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, 15(4–6), 535–547. [https://doi.org/10.1016/S0893-6080\(02\)00047-3](https://doi.org/10.1016/S0893-6080(02)00047-3)
- Juechems, K., & Summerfield, C. (2019). Where does value come from? *Trends in Cognitive Sciences*, 23(10), 836–850. <https://doi.org/10.1016/j.tics.2019.07.012>
- Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: Consensus and controversy. *Neuron*, 63(6), 733–745. <https://doi.org/10.1016/j.neuron.2009.09.003>
- Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809477>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 263–291. <https://doi.org/10.2307/1914185>
- Kennerley, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, 14(12), 1581–1589. <https://doi.org/10.1038/nn.2961>
- Kennerley, S. W., Dahmubed, A. F., Lara, A. H., & Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *Journal of Cognitive Neuroscience*, 21(6), 1162–1178. <https://doi.org/10.1162/jocn.2009.21100>
- Khaw, M. W., Glimcher, P. W., & Louie, K. (2017). Normalized value coding explains dynamic adaptation in the human valuation process. *Proceedings of the National Academy of Sciences of the United States of America*, 114(48), 12696–12701. <https://doi.org/10.1073/pnas.1715293114>
- Khaw, M. W., Li, Z., & Woodford, M. (2017). *Risk aversion as a perceptual bias* (No. w23294). National Bureau of Economic Research.
- Kimmel, D. L., Elsayed, G. F., Cunningham, J. P., & Newsome, W. T. (2020). Value and choice as separable, stable representations in orbitofrontal cortex. *Nature Communications*, 11, Article 3466. <https://doi.org/10.1038/s41467-020-17058-y>
- Klein-Flügge, M. C., Barron, H. C., Brodersen, K. H., Dolan, R. J., & Behrens, T. E. J. (2013). Segregated encoding of reward–identity and stimulus–reward associations in human orbitofrontal cortex. *The Journal of Neuroscience*, 33(7), 3202–3211. <https://doi.org/10.1523/JNEUROSCI.2532-12.2013>
- Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, 21(16), Article RC159. <https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001>
- Kobayashi, S., Pinto de Carvalho, O., & Schultz, W. (2010). Adaptation of reward sensitivity in orbitofrontal neurons. *The Journal of Neuroscience*, 30(2), 534–544. <https://doi.org/10.1523/JNEUROSCI.4009-09.2010>
- Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22(6), 1027–1038. <https://doi.org/10.1016/j.conb.2012.06.001>
- Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *The Journal of Neuroscience*, 31(14), 5504–5511. <https://doi.org/10.1523/JNEUROSCI.6316-10.2011>
- Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511618031>
- Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43, Article e1. <https://doi.org/10.1017/S0140525X1900061X>
- Loewenstein, G., Rick, S., & Cohen, J. D. (2008). Neuroeconomics. *Annual Review of Psychology*, 59, 647–672. <https://doi.org/10.1146/annurev-psyich.59.103006.093710>
- Maia, T. V. (2010). Two-factor theory, the actor–critic model, and conditioned avoidance. *Learning & Behavior*, 38(1), 50–67. <https://doi.org/10.3758/LB.38.1.50>
- Marsh, B. (2002). Do animals use heuristics? *Journal of Bioeconomics*, 4(1), 49–56. <https://doi.org/10.1023/A:1020655022163>
- Martin, R. (2011). *The St. Petersburg paradox*. Stanford Encyclopedia of Philosophy.
- Maunsell, J. H. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends in Cognitive Sciences*, 8(6), 261–265. <https://doi.org/10.1016/j.tics.2004.04.003>

- Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review*, 126(2), 292–311. <https://doi.org/10.1037/rev0000120>
- Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.
- Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, 36(2), 265–284. [https://doi.org/10.1016/S0896-6273\(02\)00974-1](https://doi.org/10.1016/S0896-6273(02)00974-1)
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3), 139–154. <https://doi.org/10.1016/j.jmp.2008.12.005>
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3), 507–520. <https://doi.org/10.1007/s00213-006-0502-4>
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669), 452–454.
- O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Annals of the New York Academy of Sciences*, 1121(1), 254–272. <https://doi.org/10.1196/annals.1401.036>
- O'Doherty, J. P. (2014). The problem with value. *Neuroscience and Biobehavioral Reviews*, 43, 259–268. <https://doi.org/10.1016/j.neubiorev.2014.03.027>
- O'Neill, M., & Schultz, W. (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron*, 68(4), 789–800. <https://doi.org/10.1016/j.neuron.2010.09.031>
- O'Reilly, R. C. (2020). Unraveling the mysteries of motivation. *Trends in Cognitive Sciences*, 24, 425–434. <https://doi.org/10.1016/j.tics.2020.03.001>
- Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *The Journal of Neuroscience*, 29(44), 14004–14014. <https://doi.org/10.1523/JNEUROSCI.3751-09.2009>
- Padoa-Schioppa, C. (2011). Neurobiology of economic choice: A good-based model. *Annual Review of Neuroscience*, 34, 333–359. <https://doi.org/10.1146/annurev-neuro-061010-113648>
- Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience*, 11(1), 95–102. <https://doi.org/10.1038/nn2020>
- Pais, D., Hogan, P. M., Schlegel, T., Franks, N. R., Leonard, N. E., & Marshall, J. A. (2013). A mechanism for value-sensitive decision-making. *PLOS ONE*, 8(9), Article e73216. <https://doi.org/10.1371/journal.pone.0073216>
- Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, 6(1), Article 8096. <https://doi.org/10.1038/ncomms9096>
- Passino, K. M., Seeley, T. D., & Visscher, P. K. (2008). Swarm cognition in honey bees. *Behavioral Ecology and Sociobiology*, 62(3), 401–414. <https://doi.org/10.1007/s00265-007-0468-1>
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, 43(1), 87–131. <https://doi.org/10.1146/annurev.ps.43.020192.000511>
- Piantadosi, S. T., & Hayden, B. Y. (2015). Utility-free heuristic models of two-option choice can mimic predictions of utility-stage models under many conditions. *Frontiers in Neuroscience*, 9, Article 105. <https://doi.org/10.3389/fnins.2015.00105>
- Pirrone, A., Azab, H., Hayden, B. Y., Stafford, T., & Marshall, J. A. (2018). Evidence for the speed–value trade-off: Human and monkey decision making is magnitude sensitive. *Decision*, 5(2), 129–142. <https://doi.org/10.1037/dec0000075>
- Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *The Journal of Neuroscience*, 27(37), 9984–9988. <https://doi.org/10.1523/JNEUROSCI.2131-07.2007>
- Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741), 233–238.
- Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556. <https://doi.org/10.1038/nrn2357>
- Raposo, D., Kaufman, M. T., & Churchland, A. K. (2014). A category-free neural population supports evolving demands during decision-making. *Nature Neuroscience*, 17(12), 1784–1792. <https://doi.org/10.1038/nn.3865>
- Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making*, 28(6), 850–865. <https://doi.org/10.1177/0272989X08327066>
- Rich, E. L., & Wallis, J. D. (2016). Decoding subjective decisions from orbitofrontal cortex. *Nature Neuroscience*, 19(7), 973–980. <https://doi.org/10.1038/nn.4320>
- Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585–590. <https://doi.org/10.1038/nature12160>
- Roesch, M. R., & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *Journal of Neurophysiology*, 90(3), 1766–1789. <https://doi.org/10.1152/jn.00019.2003>
- Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science*, 304(5668), 307–310. <https://doi.org/10.1126/science.1093223>
- Roesch, M. R., Taylor, A. R., & Schoenbaum, G. (2006). Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron*, 51(4), 509–520. <https://doi.org/10.1016/j.neuron.2006.06.027>
- Rudebeck, P. H., Saunders, R. C., Lundgren, D. A., & Murray, E. A. (2017). Specialized representations of value in the orbital and ventrolateral prefrontal cortex: Desirability versus availability of outcomes. *Neuron*, 95(5), 1208–1220.e5. <https://doi.org/10.1016/j.neuron.2017.07.042>
- Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, 70(6), 1054–1069. <https://doi.org/10.1016/j.neuron.2011.05.014>
- Salomon, T., Botvinik-Nezer, R., Gutentag, T., Gera, R., Iwanir, R., Tamir, M., & Schonberg, T. (2018). The cue-approach task as a general mechanism for long-term non-reinforced behavioral change. *Scientific Reports*, 8(1), Article 3614. <https://doi.org/10.1038/s41598-018-21774-3>
- Santos, L. R., & Rosati, A. G. (2015). The evolutionary roots of human decision making. *Annual Review of Psychology*, 66, 321–347. <https://doi.org/10.1146/annurev-psych-010814-015310>
- Schonberg, T., Bakkour, A., Hover, A. M., Mumford, J. A., Nagar, L., Perez, J., & Poldrack, R. A. (2014). Changing value through cued approach: An automatic mechanism of behavior change. *Nature Neuroscience*, 17(4), 625–630. <https://doi.org/10.1038/nn.3673>
- Schonberg, T., & Katz, L. N. (2020). A neural pathway for nonreinforced preference change. *Trends in Cognitive Sciences*, 24, 504–514. <https://doi.org/10.1016/j.tics.2020.04.002>
- Seeley, T. D. (2010). *Honeybee democracy*. Princeton University Press.
- Seeley, T. D., & Buhrman, S. C. (1999). Group decision making in swarms of honey bees. *Behavioral Ecology and Sociobiology*, 45(1), 19–31. <https://doi.org/10.1007/s002650050536>
- Seeley, T. D., Passino, K., & Visscher, K. (2006). Group decision making in honey bee swarms: When 10,000 bees go house hunting, how do they cooperatively choose their new nesting site? *American Scientist*, 94(3), 220–229. <https://doi.org/10.1511/2006.59.220>
- Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (*Apis mellifera*) and gray jays (*Perisoreus canadensis*). *Behavioral Ecology and Sociobiology*, 51(2), 180–187. <https://doi.org/10.1007/s00265-001-0420-8>
- Sharot, T., Fleming, S. M., Yu, X., Koster, R., & Dolan, R. J. (2012). Is choice-induced preference change long lasting? *Psychological Science*, 23(10), 1123–1129. <https://doi.org/10.1177/0956797612438733>

- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12), 1317–1322. <https://doi.org/10.1038/nn1150>
- Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of context-dependent choice: A new model and evidence. *PLoS Computational Biology*, 8(7), Article e1002607. <https://doi.org/10.1371/journal.pcbi.1002607>
- Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience*, 18(5), 620–627. <https://doi.org/10.1038/nn.3982>
- Stevens, J. R. (2016). Intertemporal similarity: Discounting as a last resort. *Journal of Behavioral Decision Making*, 29(1), 12–24. <https://doi.org/10.1002/bdm.1870>
- Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, 53(1), 1–26. <https://doi.org/10.1016/j.cogpsych.2005.10.003>
- Stewart, N., & Simpson, K. (2008). A decision-by-sampling account of decision under risk. *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 261–276). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199216093.003.0012>
- Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, 82(6), 1357–1366.
- Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, 304(5678), 1782–1787.
- Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, 6(5), 363–375. <https://doi.org/10.1038/nrn1666>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.
- Sutton, R. S., McAllester, D. A., Singh, S. P., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. *Advances in Neural Information Processing Systems 12*, 1057–1063.
- Takahashi, Y., Schoenbaum, G., & Niv, Y. (2008). Silencing the critics: Understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in Neuroscience*, 2, 86–99. <https://doi.org/10.3389/neuro.01.014.2008>
- Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: Violations of value invariance in human decision making. *Psychonomic Bulletin & Review*, 23(1), 22–38. <https://doi.org/10.3758/s13423-015-0858-8>
- Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729), 704–708. <https://doi.org/10.1038/19525>
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1), 31–48. <https://doi.org/10.1037/h0026750>
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281–299. <https://doi.org/10.1037/h0032955>
- Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, 102(2), 269–283. <https://doi.org/10.1037/0033-295X.102.2.269>
- Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science*, 3(6), 358–361. <https://doi.org/10.1111/j.1467-9280.1992.tb00047.x>
- Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, 15(11), 546–554. <https://doi.org/10.1016/j.tics.2011.09.008>
- Voigt, K., Murawski, C., & Bode, S. (2017). Endogenous formation of preferences: Choices systematically change willingness-to-pay for goods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(12), 1872–1882. <https://doi.org/10.1037/xlm0000415>
- Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience*, 30, 31–56. <https://doi.org/10.1146/annurev.neuro.30.051606.094334>
- Wallis, J. D., & Rich, E. L. (2011). Challenges of interpreting frontal neurons during value-based decision-making. *Frontiers in Neuroscience*, 5, Article 124. <https://doi.org/10.3389/fnins.2011.00124>
- Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, 18(6), 516–523. <https://doi.org/10.1111/j.1467-9280.2007.01932.x>
- Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4), 229–256. <https://doi.org/10.1007/BF00992696>
- Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2), 267–279. <https://doi.org/10.1016/j.neuron.2013.11.005>
- Yoo, S. B. M., & Hayden, B. Y. (2018). Economic choice as an untangling of options into actions. *Neuron*, 99(3), 434–447. <https://doi.org/10.1016/j.neuron.2018.06.038>
- Yoo, S. B. M., Sleezer, B. J., & Hayden, B. Y. (2018). Robust encoding of spatial information in orbitofrontal cortex and striatum. *Journal of Cognitive Neuroscience*, 30(6), 898–913. https://doi.org/10.1162/jocn_a_01259
- Zimmermann, J., Glimcher, P. W., & Louie, K. (2018). Multiple timescales of normalized value coding underlie adaptive choice behavior. *Nature Communications*, 9(1), Article 3206. <https://doi.org/10.1038/s41467-018-05507-8>

Received October 4, 2020

Revision received January 1, 2021

Accepted January 4, 2021 ■