Toward a Mechanistic Understanding of Human Decision Making

Contributions of Functional Neuroimaging

John P. O'Doherty and Peter Bossaerts

Computation and Neural Systems Program and Division of Humanities and Social Sciences, California Institute of Technology

ABSTRACT—*This article considers the contribution of functional neuroimaging toward understanding the computational underpinnings of human decision making. We outline the main processes likely underlying the capacity to make simple choices and describe their associated neural substrates. Relevant processes include the ability to encode a representation of the expected value or utility associated with each option in a decision problem, to learn such expectations through experience, and to modify action selection in order to choose those actions leading to the greatest reward. We provide several examples of how functional neuroimaging data have helped to shape and inform theories of decision making over and above results available from traditional behavioral measures.*

KEYWORDS—*choice; value; reward; risk; reinforcement learning; fMRI*

Whether deciding what dish to order from a restaurant menu or pondering what career path to follow, people are frequently faced with the challenge of making complex decisions, often in the absence of complete information. Decision making has long been studied at the theoretical and/or behavioral level, by economists and psychologists from both the animal-learning and cognitive-psychology traditions. These disciplines share the common goal of understanding the fundamental computations underlying simple choice. Yet such an understanding has arguably remained elusive. The modern field of decision neuroscience has flourished in recent years, aided considerably by existing theories and experimental methodologies derived from these other approaches. A driving force behind much of the current interest in applying neuroscience tools to decision making is the hope that, by studying how decisions are implemented at the neural level, it will be possible to bring clarity to a theoretical understanding that cannot be achieved through conventional behavioral studies.

In this article, we address the question of whether decision neuroscience can fulfill these expectations. We focus on functional neuroimaging, one of the principle tools available for studying brain function in humans. We argue that decision making depends on at least four distinct mechanisms: (a) encoding representations of expected future reward, (b) encoding representations of the variance in expected reward, (c) learning and updating these representations, and (d) performing actions on the basis of these representations.

Address correspondence to John P. O'Doherty, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; e-mail: jdoherty@caltech.edu.

EXPECTED REWARD, VALUE, AND UTILITY

A key concept in essentially all approaches to decision making is that choice between different options is made by considering the expected future reward associated with those options. Imaging studies in humans provide evidence that representations of expected reward are indeed encoded in the brain. The simplest way to probe expected reward is through Pavlovian conditioning, which involves passive learning of associations between an initially affectively neutral conditioned stimulus (CS) and the subsequent delivery of rewarding or punishing outcomes. In a typical study, subjects are scanned while being presented with different visual cues. One of these cues (designated the CS+) is paired repeatedly with the subsequent delivery of reward (such as a pleasant odor or a pleasant taste or

juice stimulus), whereas another cue (the CS−) is either paired with nothing or paired with an affectively neutral stimulus. Neural responses related to reward prediction can then be isolated by comparing activity to the onset of the CS+ with activity to the onset of the CS−. Such studies reveal significant CS+ related activity in regions such as the orbitofrontal cortex, amygdala, and ventral striatum (Fig. 1), implicating these areas in encoding expectations of reward (Gottfried, O'Doherty, & Dolan, 2003).

Expected-reward representations have also been investigated during performance of simple decision tasks such as the n-armed bandit task. In this task, a subject chooses among a number of different options corresponding in an abstract sense to different "arms" that might be available on a slot machine. These options pay out reward with differing probabilities or magnitudes. The subjects' objective is to establish through trial and error which arm pays the most. Imaging studies have revealed significant activity in orbitofrontal and medial prefrontal areas and the amygdala correlating with the expected reward of the chosen action during performance of such tasks (Hampton, Bossaerts, & O'Doherty, 2006).

While all theories of decision making invoke representations of expected reward, theories differ as to the precise form such representations take. Relevant properties of the reward include the magnitude or quantity of the reward expected and the probability that the reward will actually be delivered. These properties can be combined in different ways to produce variables relevant for decision making. The most fundamental combined variable is expected value (EV), which multiplies the probability of obtaining a reward with the magnitude of the reward. These concepts have been applied in human neuroimaging studies to investigate their neural correlates. Such studies have revealed neural responses to expected magnitude, probability, and EV in the ventral striatum and the medial prefrontal cortex (Tobler, O'Doherty, Dolan, & Schultz, 2007).

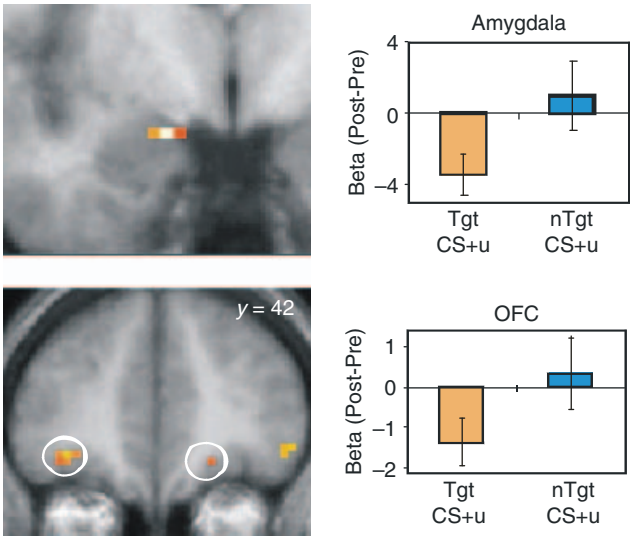

Fig. 1. Neural responses during expectation of reward. Hungry subjects were conditioned to two cues that signaled subsequent delivery of two food-related odor rewards (vanilla and peanut butter). Subjects were then fed to satiety on one of the foods, selectively devaluing the corresponding food-related odor. Subjects were then scanned again while being presented with the same two conditioned stimuli. Brain images show regions of amygdala (top) and orbitofrontal cortex (bottom; activity denoted by the colored patches) exhibiting activity in response to the cue stimuli related to the current reward value (perceived pleasantness) of the odors to which these cue stimuli have been paired. The graphs show the difference in activity in the amygdala (upper graph) and orbitofrontal cortex (OFC; bottom graph) from pre- to post satiety in response to the odor of the food fed to satiety (Tgt CS+u) compared to the response to cues associated with the odor of the food not fed to satiety (nTgt CS+u). Adapted from ''Encoding Predictive Reward Value in Human Amygdala and Orbitofrontal Cortex,'' by J.A. Gottfried, J. O'Doherty, & R.J. Dolan, 2003, *Science*, *301*, p. 1106. Copyright, 2003, American Association for the Advancement of Science. Adapted with permission.

Expected Utility

An important extension of EV is expected utility (EU): the probability of reward times its utility. EU offers an advantage over EV, by being able to account for the fact that humans and indeed other animals are sensitive not only to EV of an outcome but also to the degree of risk or uncertainty as to whether the reward will be delivered. Some individuals are risk averse, such that they prefer a safe gamble to a risky gamble if the gambles are equal in EV; other individuals may be risk neutral or even risk prone (they favor a risky compared to a nonrisky gamble with equal EV). The propensity for risk to bias choice can be accounted for in EU theory by virtue of the degree of curvature of the utility curve. For example, when given a choice between a certain $5 outcome and an uncertain $10 gamble—in which there is a 50% probability of winning the $10 but also a 50% probability of winning nothing—an individual whose utility curve is concave (because its rate of increase from $5 to $10 is much less than that from $0 to $5) will have a tendency to prefer the certain $5 outcome to the uncertain $10 gamble, because the EU of that certain outcome is greater than that of the uncertain one.

Dual Representations of Expected Value and Risk

The suggestion that individuals use a single representation of expected utility to guide choice is parsimonious, but is this actually what happens when people make decisions? An alternative possibility is that subjects use two separate signals: EV (the mean reward expected) and an additional variable that encodes the degree of risk or variance in the distribution of rewards. Neuroimaging studies can help to discriminate between these two different representational forms, even though they offer identical behavioral predictions in that both can account for risk sensitivity in behavioral choices equally well. Consistent with the mean/variance alternative, imaging studies support a distinct neural representation for risk. Several studies report increased activity in the anterior insula or lateral orbitofrontal cortex during presentation of risky as opposed to nonrisky gambles (Kuhnen & Knutson, 2005). These areas are often implicated in aversive processing such as responding to pain or its anticipation, as well as

to monetary loss; thus responses to risk in these areas could reflect the negative affective state it engenders. Risk signals have also been found in the firing patterns of dopamine neurons corresponding to a ramping (or monotonically increasing) activity in the interval during which a reward outcome is being anticipated (Fiorillo, Tobler, & Schultz, 2003). Similar signals have been found in functional magnetic resonance imaging (fMRI) studies in target areas of dopamine neurons such as the ventral striatum (Preuschoff, Bossaerts, & Quartz, 2006; Fig. 2).

These findings suggest that both EV and risk are indeed encoded separately, arguing against the possibility that the brain exclusively encodes representations of EU. Whether a unified representation of EU is present in the brain at all remains an open question. To our knowledge, conclusive evidence for the existence of such a signal at the neural level (distinct from EV) has yet to be reported. It should also be noted that EU theory has long been challenged on the basis of behavioral evidence. For example, subjects tend to be risk seeking for gambles involving losses but risk averse for gambles involving gains. Such anomalies have led to proposed variations in EU such as prospect theory (Kahneman & Tversky, 1979). Preliminary evidence that neural activity during decision making also reflects such asymmetries has emerged (De Martino, Kumaran, Seymour, & Dolan, 2006; Tom, Fox, Trepel, & Poldrack, 2007).

LEARNING EXPECTED-REWARD SIGNALS

The finding of EV signals in the brain raises the question of how such signals are learned in the first place. An influential theory from behavioral psychology put forward by Rescorla and Wagner suggests that learning of reward expectancy is mediated by the degree of surprise engendered when an outcome is presented—or more precisely, the difference between what is expected and what is received (Rescorla & Wagner, 1972). Formally this is called a prediction error, which can take either a positive or negative sign depending on whether an outcome is greater than expected (which would lead to a positive error signal) or less than expected (which would lead to a negative error signal). This prediction error is used to update the expected reward associated with a particular stimulus or cue in the environment. Prediction-error signals in the brain were first uncovered through observation of the firing patterns of dopamine neurons in nonhuman primates during performance of simple reward-conditioning tasks. These neurons were found to demonstrate a firing pattern that closely resembles the prediction-error signals present in temporal-difference learning (Sutton, 1988)—an extension of the Recorla-Wagner model that captures how reward predictions are formed at different time points within a trial rather than merely on a trial-by-trial basis. fMRI studies in humans have found evidence for prediction-error signals in target regions of dopamine neurons such as the ventral striatum and orbitofrontal cortex (O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003).

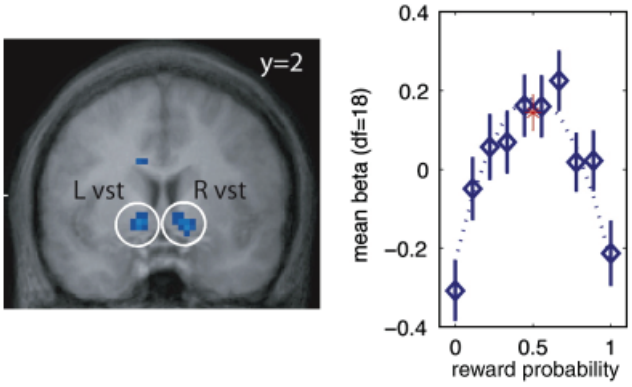

Fig. 2. Anticipatory risk signals in the human brain. The brain image (A) shows ramping (monotonically increasing) anticipatory signals in the left and right ventral striatum (vst) related to the risk or variance in anticipated reward. The plot (B) shows fMRI responses in this region as a function of reward probability, demonstrating a maximal response on trials when reward delivery is least predictable (i.e., with a probability $p = 0.5$), and a minimal response when the reward delivery is most predictable and is either certain to occur or not to occur ($p = 0$ and $p = 1$, respectively). The red star shows activation levels during a different anticipation phase in the trial when risk is always maximal (i.e. when $p = 0.5$). Adapted from "Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures," by K. Preuschoff, P. Bossaerts, & S.R. Quartz, 2006, *Neuron*, *51*, p. 384. Copyright 2006, Elsevier. Adapted with permission.

Analysis of the precise neural mechanisms underlying learning of reward predictions in the context of the Rescorla-Wagner model (and its extensions) is still ongoing. The model relies on a crucial parameter that determines to what extent prediction errors influence future predictions. This parameter should change with the uncertainty in the environment (Behrens, Woolrich, Walton, & Rushworth, 2007; Preuschoff & Bossaerts, 2007; Yu & Dayan, 2003). Implementation of such changes presupposes, among other things, that risk is encoded: If risk is expected to be high, then prediction errors ought not to change future predictions much—they are predicted to be sizeable anyway. But note that we already discussed encoding of risk in the context of choice. This suggests that such encoding may play a dual role—namely, to guide choice, and to modulate learning.

ACTION SELECTION FOR REWARD

Once predictions of future reward are made, the next step is to use these predictions to guide behavior. More specifically, individuals need to bias their action selection to choose those actions leading to the greatest probability of obtaining future reward or avoiding punishment.

Reinforcement-Learning Models

Useful insights into how humans or other animals might solve this problem has come from a branch of computer science known as reinforcement learning (RL; Sutton & Barto, 1998). According to RL, in order to choose optimally between different actions, an agent needs to maintain internal representations of the

expected reward available on each action and, subsequently, choose the action with the highest future reward. Also central to RL algorithms is a prediction-error signal, which is used to learn and update EVs for each action through experience, just as in the Rescorla-Wagner model. In one such RL model, called the Actor/Critic, action selection is conceived as involving two distinct components: a critic, which learns to predict future reward associated with particular states in the environment, and an actor, which chooses specific actions in order to move the agent from state to state according to a learned policy. Evidence from neuroimaging studies supports the presence of an actor/critic-like architecture in the brain, even though a number of alternative RL architectures (such as Q-learning, in which action values are learned directly and not indirectly via a separate critic) can capture human choice behavior equally well. In one study, the dorsal striatum was found to be specifically engaged when subjects were actively performing instrumental responses in order to obtain reward—consistent with a role for this region in implementing the actor—whereas the ventral striatum was found to be active in both instrumental and Pavlovian tasks—consistent with a role for that region in the critic (O'Doherty et al., 2004). These results suggest a dorsal–ventral distinction within the striatum whereby the ventral striatum is more concerned with Pavlovian or stimulus–outcome learning while the dorsal striatum is more engaged during learning of stimulus-response or stimulus–response–outcome associations.

Limitations of RL Models

When taken together with a host of other findings from humans and other animals, the data we have discussed suggest that RL models provide a good account for how people learn to make choices through experience. Still, such models fail to account for all aspects of human choice behavior. Simple RL models assume that information gained about the rewards available from choosing one action provides no information about the rewards available from choosing an alternative action. Yet, in many situations, interdependencies between different actions do exist, and if subjects can exploit these, greater reward will ensue. One of the simplest examples of a decision task with such an abstract rule is probabilistic reversal learning. In essence, this is a 2-armed bandit problem in which arms are interchanged at times. The structure in this task is the negative correlation between the rewards available on the two arms: When choice of one arm is "good," the other is "bad," and vice versa—but after a time, the contingencies reverse. Evidence from neuroimaging indicates that EV representations in the ventromedial prefrontal cortex incorporate such a task structure, manifested by an inference-mediated change in expectations following a reversal of stimulus choice (Hampton et al., 2006). These results suggest that human subjects do indeed incorporate knowledge of interdependencies when making decisions, providing empirical evidence to support the application of an important extension of reinforcement learning to human decision making.

OUTSTANDING ISSUES

Although decision neuroscience and the related area of neuroeconomics are arguably still in their infancy, we have here highlighted several cases in which neuroimaging findings have informed a theoretical understanding of decision making that could not be achieved through behavioral methods alone. Many questions remain unanswered, and space precludes an exhaustive consideration of these here. One important issue is whether different neural systems are engaged when explicit information about the properties of the gamble, such as the probabilities and magnitudes, is provided, compared to when the relevant information is learned through experience in the absence of explicit information. Another open question is how the chosen action comes to be chosen in the first place. Although imaging studies provide evidence that the brain encodes expectations of reward associated with the chosen action, no study has yet distinguished between regions actively determining the decision itself from those merely reporting its consequences.

Recommended Reading

Daw, N.D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*, 199–204. A review of the insights into reward learning and decision making gained from applying formal computational models of learning to neural data.

Knutson, B., & Cooper, J.C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*, 411–417. Provides an overview of recent functional neuroimaging studies on reward prediction and details the specific neural structures implicated in this function.

Montague, P.R., & Berns, G.S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, *36*, 265–284. Provides an accessible overview on the relevance of neural data for understanding valuation processes in humans and other animals.

Acknowledgments—This work was supported by a Searle Scholarship and a grant from the Gordon and Betty Moore foundation to JOD, and by funds from the Swiss Finance Institute to PB.

REFERENCES

Behrens, T.E., Woolrich, M.W., Walton, M.E., & Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214–1221.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R.J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*, 684–687.

Fiorillo, C.D., Tobler, P.N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*, 1898–1902.

Gottfried, J.A., O'Doherty, J., & Dolan, R.J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*, 1104–1107.

Hampton, A.N., Bossaerts, P., & O'Doherty, J.P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience*, *26*, 8360–8367.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–292.

Kuhnen, C.M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*, 763–770.

O'Doherty, J., Dayan, P., Friston, K., Critchley, H., & Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*, 329–337.

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*, 452–454.

Preuschoff, K., & Bossaerts, P. (2007). Adding prediction risk to the theory of reward learning. *Annals of the New York Academy of Sciences*, *1104*, 135–146.

Preuschoff, K., Bossaerts, P., & Quartz, S.R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, *51*, 381–390.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Crofts.

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*, 9–44.

Sutton, R.S., & Barto, A.G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Tobler, P.N., O'Doherty, J.P., Dolan, R.J., & Schultz, W. (2007). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *Journal of Neurophysiology*, *97*, 1621–1632.

Tom, S.M., Fox, C.R., Trepel, C., & Poldrack, R.A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Yu, A.J., & Dayan, P. (2003). Expected and unexpected uncertainty. *Advances in Neural Information Processing Systems*, *15*, 157–164.