8

Reward:

NEURAL CIRCUITRY FOR SOCIAL VALUATION

Brian Knutson and G. Elliott Wimmer

During an Enlightenment-era correspondence with Pierre de Fermat, Blaise Pascal proposed that individuals compute "expected value" (EV) as the product of the expected magnitude and probability of a potentially favorable gamble. Since then, the notion of EV has played a pivotal role in both psychological (Bandura, 1977; Rotter, 1972) and economic theory (von Neumann & Morgenstern, 1944). Although theorists sometimes assume that people will behave in ways that will maximize EV, empirical research has documented exceptions (Kahneman & Tversky, 1979). Nonetheless, anticipated gain magnitude and probability provide useful anchors for attempts to understand how people process rewards. If the brain computes EV prior to a potentially rewarding outcome, then scientists can elicit and study the representation of EV. In this overview, we describe neuroimaging research designed to elucidate how the brain represents EV and then consider emerging evidence that neural representation of EV extends to, pervades, and may even influence social exchange.

BACKGROUND

If the amount of effort an animal expends to obtain a stimulus indexes value, then self-stimulation represents one of the most extreme examples of valuation. Social psychologist James Olds and physicist Peter Milner serendipitously discovered self-stimulation in 1954, while attempting to electrically stimulate arousal centers in the midbrain of rats (Olds &

Milner, 1954). Olds and Milner noticed that a rat that had had an electrode erroneously implanted near the nucleus accumbens (NAcc) rather than in arousal centers of the midbrain not only showed energized behavior when stimulated but also spontaneously returned to the corner of a table where it had been stimulated the day before. After devising an apparatus that allowed the rat to self-administer stimulation by pressing a bar, Olds and Milner found that the rat worked vigorously to do so, to the point of exhaustion and the exclusion of all other activities (e.g., eating, drinking, sex, and sleep). Since its discovery, self-stimulation behavior has been demonstrated in all other mammalian species studied (Olds & Fobes, 1981), including humans (Bishop, Elder, & Heath, 1963).

Brain sites that support self-stimulation ascend from deep in the midbrain to higher subcortical regions (i.e., lateral hypothalamus, medial amygdala, and ventral striatum). Some cortical regions also support self-stimulation but do so less robustly (i.e., orbitofrontal cortex and mesial prefrontal cortex). Subsequent innovations in histochemical mapping indicated that the neurotransmitter dopamine could be found in many of these regions (Falck & Hillarp, 1959). Regions deep in the midbrain appeared to house the bodies of dopamine neurons, which projected to the subcortical and cortical regions. Microinjection studies later indicated that rats would expend similar effort to self-administer dopamine-like chemicals to many of these sites (McBride, Murphy, & Ikemoto, 1999). Methods for visualizing synaptic activity on a subsecond time scale revealed dopamine release in subcortical and cortical projection areas as rats anticipated rewards ranging from food to sex to commonly abused drugs (Wightman & Robinson, 2002). Finally, electrophysiological recording of midbrain dopamine neurons in monkeys indicated that even after learning had stabilized, dopamine neurons continued to fire during anticipation of rewards but transiently ceased firing when anticipated rewards were not delivered (Schultz, Dayan, & Montague, 1997). Together, this remarkable progression of findings implicates mesolimbic dopamine projections in both self-stimulation behavior and reward anticipation and, by extension, in the computation of EV.

Most self-stimulation studies have been conducted with nonhuman subjects due to the invasiveness of implanting electrodes in the brain. However, technology for visualizing human brain activity became available near the end of the 20th century. For instance, positron emission tomography (PET) enabled visualization of local neural utilization of oxygen, glucose, and even certain neurotransmitters. In addition to demonstrating that dopamine is released in the ventral striatum when people play engaging games (Koepp et al., 1998; Pappata et al., 2002; Zald et al., 2004), PET researchers have demonstrated that dopamine release in the ventral striatum caused by amphetamine injection correlates with self-reported positive arousal (or euphoria) but not with negative arousal (or fear; Drevets et al., 2001; Mawlawi et al., 2001; Volkow et al., 1999). However,

although PET can give researchers clues about what neurotransmitters are released, its temporal resolution (i.e., approximately 120 seconds per brain scan) limits inferences about when release occurs. On the other hand, although event-related functional magnetic resonance imaging (fMRI) provides information only about local changes in oxygenation (hereafter "activation"), it does provide adequate temporal resolution (i.e., about 1–2 seconds per brain scan) for researchers to infer when activation occurs. Thus the remainder of this overview focuses primarily on rapidly emerging findings from event-related fMRI.

MONETARY REWARD

Determination of the neural basis of valuation in humans constitutes one of the most basic challenges presently confronting affective neuroscience (Davidson & Sutton, 1995; Panksepp, 1991). Thus our laboratory initially used event-related fMRI to attempt to identify brain regions that represent components of EV (i.e., the magnitude and probability of expected gains). As exemplified by rapid advances in vision research (Engel et al., 1994), programmatic functional brain mapping research often progresses through stages that include (1) visualization of relevant brain regions; (2) parametric manipulation of activation in those regions; and (3) exploration of alternative functional hypotheses for activation in those regions. Thus an initial challenge was to visualize activation in the mesolimbic pathway in general and in the ventral striatum in particular. Although event-related fMRI provides adequate spatiotemporal resolution for visualizing second-to-second activation changes in small subcortical regions, we faced the additional challenge of identifying compelling incentives. We adopted monetary incentives because they are widely valued (i.e., most people will work for money), can carry either positive or negative value (i.e., can be gained or lost), and can be scaled to different magnitudes (and thus parameterized). Inspired by the early research of Pavlov with dogs (Pavlov, 1927) and more recent work of Schultz with monkeys (Schultz et al., 1997), we designed a "monetary incentive delay" (MID) task for use in humans undergoing fMRI (Knutson, Westdorp, Kaiser, & Hommer, 2000).

Honoring a traditional ethological distinction between appetitive and consummatory behavior (Craig, 1918), the MID task is designed to evoke both anticipation of and reactions to monetary gain and loss (see Figure 8.1). A typical MID task trial includes four components: (1) viewing a cue (cue; 250–2,000 milliseconds); (2) waiting (anticipation; 2,000–3,000 milliseconds); (3) responding to a rapidly presented target with a button press (target; 160–350 milliseconds); (4) receipt of trial-based and cumulative feedback about gain or loss (outcome; 2,000 milliseconds; Knutson, Fong, Bennett, Adams, & Hommer, 2003). On all trials, participants are instructed to respond as rapidly as possible when targets appear, with the

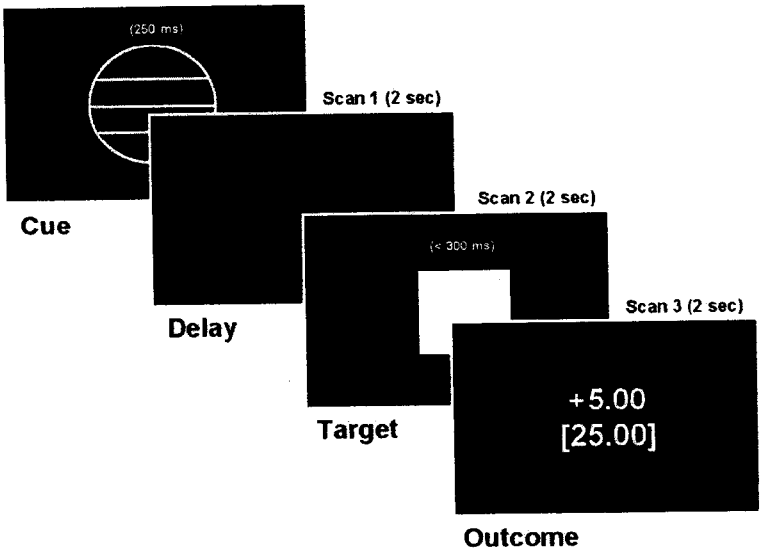

FIGURE 8.1. Monetary incentive delay task trial structure. Adapted from Knutson, Fong, Bennett, Adams, and Hammer (2003). Copyright 2003 by TK. Adapted by permission.

goal of pressing the button before targets disappear. Cue features (e.g., shape, horizontal or vertical lines) indicate whether participants can acquire gains or avoid losses by responding to the subsequently presented targets. Because of the separation of anticipation and outcome periods in the task, investigators can infer which brain regions were recruited by different stages of incentive processing.

The MID task is designed to change affect rather than overt behavior. Behavioral performance can be controlled across incentive conditions by varying the range of target speeds, which are typically set to a challenging but not impossible level of difficulty (e.g., to elicit a hit rate of 66%). However, different conditions reliably elicit distinct affective experiences. Specifically, presentation of gain cues primarily elicits positive arousal (e.g., "excitement"), whereas presentation of loss cues primarily elicits negative arousal (e.g., "anxiety") proportional to cue magnitude (Knutson et al., 2003). Changes in anticipatory affect have been verified using both retrospective and online ratings (Knutson, Nielsen, Larkin, & Carstensen, 2005). Interestingly, both gain and loss anticipation and outcomes elicit changes in valence, but whereas anticipation also elicits increased arousal, outcomes do not.

Combined with event-related fMRI, the MID task has yielded novel insights about the dynamics of human reward processing. In the first pub-

lished fMRI study to manipulate monetary incentives, we observed striatal and mesial frontal activation when participants engaged in incentive trials (both gain and loss) versus nonincentive trials over the course of the entire trial (Knutson et al., 2000). However, these activations occurred in regions that were located more dorsal than the mesolimbic pathway, suggesting that whole-trial comparisons did not afford adequate temporal resolution for resolving more rapid anticipatory changes in activation. Indeed, in a second study that utilized several different magnitudes of incentives in which modeling focused on anticipation only, the more ventral NAcc was preferentially activated by gain anticipation but not loss anticipation (Knutson, Adams, Fong, & Hommer, 2001). A third study in which anticipation and outcome were analyzed separately replicated this finding and further indicated that gain outcomes instead activated the mesial prefrontal cortex (MPFC; Knutson, Fong, Adams, Varner, & Hommer, 2001). Whereas the magnitude of anticipated gain was manipulated in these studies, the probability of anticipated gain was held constant (i.e., approximately 66% probability of obtaining gain or avoiding loss in a given trial).

In a subsequent study, we manipulated the probability, as well as the magnitude, of anticipated gain, thus independently altering both components of EV. Whereas NAcc activation was proportional to the magnitude of anticipated gains as in prior studies, MPFC activation was sensitive to the probability of anticipated gains (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005). Together, these results not only verify the involvement of mesolimbic circuitry in the computation of both components of EV (see Figure 8.2) but also further imply that whereas NAcc activation increases with anticipated gain magnitude, MPFC activation increases with anticipated gain probability. Theoretically, the findings suggest a possible mechanism for peoples' insensitivity to probability during risk assessment (Kahneman & Tversky, 1979), as cortical representation of probability may require more effortful processing than subcortical representation of magnitude. Clinically, the findings suggest that patients with cortical lesions of the MPFC (e.g., the historic case of Phineas Gage) may be able to anticipate gain magnitude but may be less able to adjust expectations according to gain probability (Camille et al., 2004; Knutson & Cooper, 2005).

Together, these findings demonstrate NAcc activation during gain anticipation and verify that this activation increases proportional to anticipated gain magnitude. Comparison of the gain anticipation hypothesis with alternative accounts of NAcc activation is ongoing but presently incomplete. The design of the MID task can address some prominent alternative hypotheses. According to one account, the NAcc activates in response to surprising or unpredicted stimuli (Berns, McClure, Pagnoni, & Montague, 2001), which implies that activation should occur in response to all incentive outcomes. However, in the MID task, the most robust NAcc activation occurs during gain anticipation rather than in response to gain outcomes. A

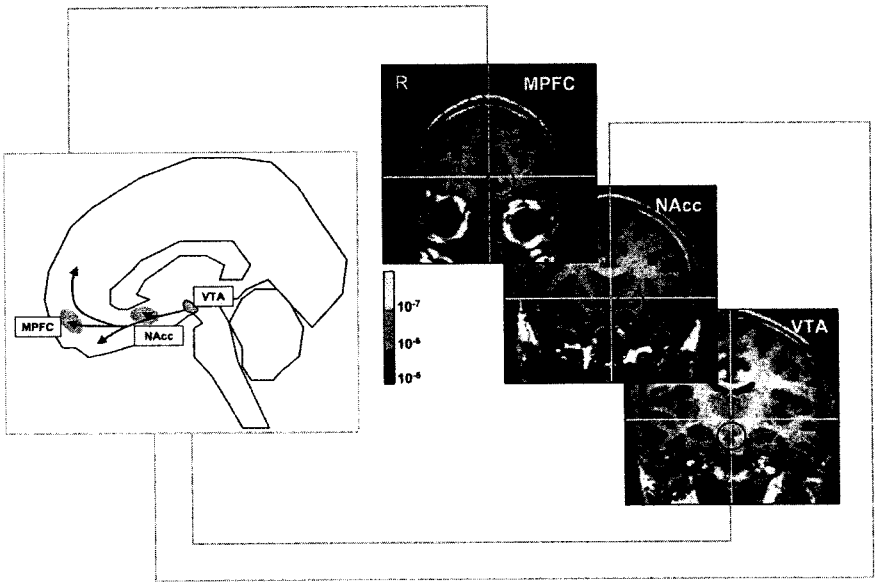

FIGURE 8.2. Activation in mesolimbic volumes of interest correlate with a linear model of expected value (i.e., anticipated gain magnitude X probability. MPFC; mesial prefrontal cortex; NAcc, nucleus accumbens; VTA, ventral tegmental area; Adapted from Knutson, Taylor, Kaufman, Peterson, and Glover (2005). Copyright 2005 by TK. Adapted by permission.

Lc/

acquired from Society for Neuroscience

second account posits that the NAcc activates during anticipation of any arousing or salient event (positive or negative; Berridge & Robinson, 1998), which implies that the NAcc should show similar increases in activation during gain or loss anticipation, as subjects report experiencing similar levels of arousal when anticipating gains and losses. However, we have repeatedly and consistently observed greater NAcc activation during anticipation of gains than losses. A third account posits that NAcc activation facilitates motor preparation (Mogenson, Jones, & Yim, 1980), which again implies that the NAcc should show similar activation during anticipation of gains and losses, as participants respond to obtain gains and avoid losses with similar speed. However, again, the NAcc shows greater activation during anticipation of gains than during anticipation of losses. A fourth account might posit that NAcc activation should occur in the context of learning and thus should show less activation after learning has stabilized (Dickinson, 1994). However, even after implicit and explicit learning have stabilized, we continue to observe prominent NAcc activation during anticipation of gains. Various combinations of these hypotheses remain to be tested, but at present, the gain anticipation hypothesis best fits

the pattern of NAcc activation observed in several experiments. But it may be associated positive arousal rather than goal representation that best correlates with NAcc activation, as individual differences in positive arousal correlate with individual differences in NAcc activation to large potential gains of the same magnitude (Bjork et al., 2004; Knutson, Adams, et al., 2001; Knutson, Taylor, et al., 2005).

Although the preceding overview focuses on the work of our laboratory, many other investigators have successfully used monetary rewards to elicit mesolimbic activation with fMRI, starting with initial demonstrations (Delgado, Nystrom, Fissell, Noll, & Fiez, 2000; Elliott, Friston, & Dolan, 2000) and followed by increasingly sophisticated paradigms that have been well summarized elsewhere (Knutson & Cooper, 2005; McClure, York, & Montague, 2004; O'Doherty, 2004). Other researchers have independently replicated the finding that anticipation of gains elicits greater mesolimbic activation than anticipation of losses (Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001) and that anticipation of gains elicits increased activation in the NAcc whereas gain outcomes elicit activation of the MPFC (Wittmann et al., 2005), even in noncontingent tasks that do not require a motor response (Ramnani, Elliott, Athwal, & Passingham, 2004).

OTHER REWARDS

fMRI researchers have also discovered that other rewards elicit activity in mesolimbic circuits, including pleasant tastes (Berns et al., 2001; O'Doherty, Rolls, Francis, Bowtell, & McGlone, 2001), pleasant smells (Anderson et al., 2003; Gottfried, O'Doherty, & Dolan, 2003), pleasant touch (Rolls et al., 2003), and pleasant sounds (such as music; Menon & Levitin, 2005). Importantly, all of these researchers empirically demonstrated that participants judged stimuli to be pleasant relative to neutral or unpleasant stimuli. However, most of these studies focused on brain responses to stimulus delivery and did not control for anticipation. One study distinguished anticipation from outcome in the case of cued delivery of pleasant, neutral, and unpleasant tastes (O'Doherty, Deichmann, Critchley, & Dolan, 2002). The investigators reported activation of midbrain, ventral striatum, and orbitofrontal cortex (OFC) during anticipation of pleasant taste but only of OFC in response to pleasant taste receipt or outcome. Though more work remains to be done, these findings are consistent with the notion that NAcc activation during gain anticipation generalizes to nonmonetary rewards. Notably, although rewarding stimuli elicit mesolimbic activation, these activations do not critically appear to depend on whether rewards are unlearned (a.k.a. "primary"), such as pleasant taste, or learned (a.k.a. "secondary"), such as money.

If NAcc activation occurs during gain anticipation, then it might bias subsequent cognition and behavior in ways that can promote the seeking of

gains (Ikemoto & Panksepp, 1999). With respect to cognition, we have recently demonstrated that monetary reward cues can enhance memory for subsequently presented scenes and that this memory enhancement effect depends on the extent to which cues elicit NAcc and midbrain activation in individual participants (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006). With respect to behavior, because risk involves weighing potential gains against potential losses, one might hypothesize that an increase in gain anticipation would promote risky choices, whereas an increase in loss anticipation would instead promote riskless choices. Using a financial trading task in combination with fMRI, we recently demonstrated that anticipatory NAcc activation predicts switching to a risk-seeking strategy (i.e., choosing stocks rather than bonds), whereas anterior insula activation predicts the opposite switch to a risk-avoidant strategy (i.e., choosing bonds rather than stocks; Kuhnen & Knutson, 2005). Thus emerging evidence is beginning to suggest that anticipatory NAcc activation may modulate subsequent cognition and behavior in ways that promote gain seeking.

Up to this point, we have implicated mesolimbic circuitry in general, and the NAcc in particular, in anticipation of various rewards and also possibly in modulating subsequent behavior. Many questions still remain to be answered. For instance, fMRI visualization of neural activation related to gain anticipation raises the question of whether a distinct circuit related to loss anticipation can be visualized (e.g., the anterior insula; Kuhnen & Knutson, 2005). Another set of questions regards whether these findings will generalize to the realm of social interaction. Specifically, can socially rewarding stimuli also activate mesolimbic circuitry, and can activation of these regions also influence subsequent social behavior?

SOCIAL REWARD

As with the broader spectrum of rewards, social rewards can be specifically defined as attributes or behavior of others that an organism will expend effort to obtain and that can either be unlearned or learned. Even in the case of novel social stimuli such as faces, people prefer symmetrical to asymmetrical structures and smiling to frowning expressions (Grammer & Thornhill, 1994; Knutson, 1996). However, people can also rapidly and flexibly learn to assign reward value to novel social stimuli, with lasting consequences. In this section, we examine whether unlearned and learned socially rewarding stimuli elicit mesolimbic activation and whether mesolimbic activation can influence subsequent social behavior (Adolphs, 2001).

Following initial reports of amygdalar activation to fearful faces (Breiter et al., 1996), some of the first fMRI studies to focus on social incentives utilized facial stimuli. An initial study reported ventral striatal activation to forward-gazing faces that participants rated as attractive

(Kampe, Frith, Dolan, & Frith, 2001). However, as noted in a later correction (Kampe, Frith, Dolan, & Frith, 2002), activation foci were located in the thalamus rather than the ventral striatum, as initially reported. A subsequent study reported mesolimbic activation in general and NAcc activation in particular in male participants exposed to female faces (Aharon et al., 2001). NAcc activation correlated both with rated attractiveness and with the number of button presses participants made to continue viewing the female faces in a separate behavioral experiment. Both of these studies had block designs and so may have included activation that occurred during both anticipation and outcome. To control for anticipatory confounds, a third study used an event-related design and reported that both male and female participants showed increased activation of the mPFC and OFC as a function of the face's participant-rated attractiveness and whether the face was judged as happy or not (O'Doherty et al., 2003). Together, these findings suggest that both the attractiveness and the happiness of novel faces can activate the MPFC, and they further raise the possibility that anticipation of viewing attractive faces may activate the NAcc, though the last implication has not been directly tested.

For some, other novel social rewards may include erotic stimuli (Lang, Greenwald, Bradley, & Hamm, 1993). A number of fMRI studies have investigated neural correlates of exposure to erotic visual stimuli. Initial studies primarily contrasted erotic films with nonerotic films among heterosexual males and reported widespread activations of subcortical and cortical regions, including but not limited to mesolimbic regions (Arnow et al., 2002; Garavan et al., 2000; Karama et al., 2002; Mouras et al., 2003; Park et al., 2001). Unlike studies of other rewarding social stimuli (e.g., faces), studies of erotic films may produce more widespread activation because investigators have modeled responses to lengthy and dynamic stimuli rather than to discrete static images. Accordingly, when investigators concurrently measured online affective indices such as penile tumescence during viewing of erotic films and correlated these with brain activation, they found more focused mesolimbic and mesial cortical patterns of activation (Arnow et al., 2002; Ferretti et al., 2005). Further, subsequent studies of more temporally constrained erotic pictures clearly demonstrate ventral striatal activation, as well as increased activation of visual processing pathways, in both male and female participants, who reported similar levels of sexual motivation (Hamann, Herman, Nolan, & Wallen, 2004).

Social reward can also be learned, as in the case of bonding. Prototypical examples include the nurturing attachment that develops between mother and infant, as well as the romantic attachment that develops between lovers. Accordingly, fMRI experiments have begun to investigate neural responses to infants and lovers. With respect to infants, one study reported lateral OFC activation when mothers viewed their own versus other infants, and this activation was correlated with positive mood when viewing the pictures (Nitschke et al., 2004). However, this study suffered

from signal loss in the mPFC and the striatum. A second study without signal loss in these regions similarly reported activation of lateral OFC but also of the striatum (including ventral striatum) when mothers viewed pictures of their own versus other infants (Bartels & Zeki, 2004). These studies suggest that viewing one's own infant can activate mesolimbic circuitry. FMRI researchers have also investigated exposure to pictures of lovers. An initial study reported that viewing pictures of lovers versus friends activated striatal regions, as well as the anterior insula and anterior cingulate (Bartels & Zeki, 2000), and these results were replicated in a follow-up study (Bartels & Zeki, 2004). A third study of recent lovers (i.e., relationships begun less than 3 months prior to the study) also found that viewing lovers versus acquaintances activated striatal regions and midbrain nuclei, but not the other cortical regions (Aron et al., 2005). Thus studies suggest that viewing pictures of lovers can activate striatal and midbrain regions of the mesolimbic circuit.

In addition to innate bonding mechanisms, social reward can be learned over the course of repeated interactions. For instance, all human cultures value reciprocity, and repeated reciprocity represents an evolutionarily stable strategy that can sustain cooperation, even among strangers (Trivers, 1971). Behavioral economists have devised ingenious games to elicit reciprocity (or not), such as the prisoner's dilemma game (Axelrod & Hamilton, 1981). In an iterated version of the prisoner's dilemma, two players repeatedly and independently choose either to cooperate or defect. If they mutually cooperate, both win, and if they mutually defect, both lose. However, if one player defects and the other cooperates, the defector wins more than if he or she had cooperated, whereas the cooperator loses more than if he or she had defected. The standard in behavioral economics is for partners to play with actual money, and all of the studies described here adhere to this standard.

In an initial fMRI study of the prisoner's dilemma game, investigators found that mutually cooperative outcomes (rated as most desirable by the female participants) elicited more mesolimbic activation (specifically, in the NAcc, caudate, mPFC, and anterior cingulate) than did outcomes elicited by other strategies. Further, peak activation in the NAcc, but not other regions, correlated with a tendency to engage in repeated mutual cooperation (Rilling et al., 2002). Similar but less robust patterns of mesolimbic activation were observed when participants played against a computer. These researchers replicated the same pattern of findings in a follow-up study that utilized a one-shot version of the prisoner's dilemma game with both male and female participants. After participants chose to cooperate, activation in NAcc and mPFC increased when partners also cooperated but decreased when partners chose to defect. In another study, participants first played a prisoner's dilemma game outside the scanner with partners (identified by pictures) who consistently responded to the participant's cooperation with either cooperation or defection (Singer, Kiebel, Winston, Dolan,

& Frith, 2004). In a subsequent fMRI session, participants showed increased medial amygdala, putamen/NAcc, and left insula activation while viewing pictures of intentional cooperators versus nonplayers. Thus social stimuli arbitrarily associated with prior cooperation also can elicit mesolimbic activation.

Another game that elicits social reciprocity is the trust game, in which an investor receives money and then decides how much to either invest in a trustee or retain. The invested amount then triples, and the trustee must decide how much to pay back the investor or retain (Berg, Dickhaut, & McCabe, 1995). In an fMRI study, pairs of participants were scanned simultaneously while playing repeated trust games (King-Casas et al., 2005). When trustees learned that the investor had increased their investment relative to previous investments, activation increased in the head of the caudate/NAcc. As the task progressed and investments increased, the onset of this activation in the trustees' brain began earlier, until it preceded revelation of increased investments.

However, social reciprocation need not always involve cooperation. In some cases, people will expend inordinate amounts of effort and time to reciprocate defection, even at the risk of substantial personal loss (e.g., the case of revenge). Using PET imaging rather than fMRI, researchers investigated brain activation associated with anticipation of revenge in the context of a trust game with male participants (de Quervain et al., 2004). The researchers gave money to participants, who then invested in trustees. After the investment tripled, trustees then either cooperated by returning part of the investment or defected by returning nothing. Investors could then decide whether or not to punish defectors (by taking away their money), during which time their brains were scanned with PET. Participants who chose to punish defectors showed activation of the head of the caudate (just above the NAcc). The investigators inferred that this activation was related to the desire to punish defectors, as activation in the head of the caudate predicted how much money participants were willing to spend to punish defectors in a separate condition.

Other researchers scanned participants with fMRI as they played a more complicated matrix game for money, in which participants attempted to coordinate choices with a partner outside the scanner to maximize earnings. Some choices maximized only the participant's earnings or only the partner's earnings, whereas others maximized the pair's joint earnings. Participants had to take the perspectives of their partners and imagine their partners taking their own perspectives in order to identify "equilibrium" solutions that would maximally benefit both partners. While undergoing fMRI, participants either simultaneously chose with the partner or simply guessed their partner's strategy. The only brain area that showed significant activation when participants coordinated to choose equilibrium solutions versus merely identifying partners' strategies was the NAcc (Bhatt & Camerer, 2005).

As with research utilizing other types of rewards, the collective findings of these social game studies are consistent with the notion that NAcc activation indexes gain anticipation (Knutson, Adams, et al., 2001). Gain anticipation may occur in the context of reciprocated cooperation, but it also may occur prior to reciprocated defection (e.g., in the case of anticipated revenge; Knutson, 2004), and it seems related to social coordination in all cases. Although the findings suggest that social rewards activate qualitatively similar regions to other types of rewards, they also raise the possibility that social rewards may have a more pronounced quantitative effect on mesolimbic activation than monetarily equivalent nonsocial rewards. Of course, social coordination may additionally engage qualititatively distinct brain areas associated with other functions (not reviewed here), such as regions implicated in perspective taking, which lie along the mesial wall of the prefrontal cortex (Gallagher & Frith, 2003; McCabe, Houser, Ryan, Smith, & Trouard, 2001).

In addition to converging with literature on other rewards, neuroimaging research on social rewards brings new light to bear on how neuroscience methods can inform the study of social behavior. Although social behavior is complex, prediction of complex phenomena does not necessarily require complex theory. The dynamic interplay of simple mechanisms might also generate complex behavior (Braitenberg, 1984). For example, neuroimaging evidence suggests that a process as simple as reward anticipation (and accompanying affect) may underlie a diverse range of social phenomena, including cooperation, revenge, moral attribution, and strategic coordination. Thus neuroimaging may inform an understanding of social behavior by helping researchers to deconstruct complex processes and to winnow out unnecessarily complex theories based on questionable assumptions. Social behavior can be studied empirically, incrementally, hierarchically, and systematically from a neuroscience perspective.

IMPLICATIONS

Humans are not just information processors. Humans are also value processors. In fact, value may take precedence over other types of information (Zajonc, 1980), because organisms that cannot efficiently assess value may not survive long enough to represent their genes in future generations (Panksepp, Knutson, & Burgdorf, 2002). Because they cannot survive or reproduce alone, social connections are among the most highly valued of incentives for mammals (MacLean, 1990). Thus mammalian value computation should not only reside deep in the processing hierarchy but should also be especially attuned to social incentives. To survive and procreate, mammals should not only be able to assess value reactively but also proactively. Indeed, one reason for the expansion of the forebrain in primates (Semendeferi, Lu, Schenker, & Damasio, 2002) may involve enhanc-

ing the ability to assign value to future events, particularly those related to social interaction.

Emerging techniques that allow researchers to visualize neural activity at the speed of phenomenology will revolutionize psychology. Part of this revolution will be methodological. Researchers can now track second-to-second changes in small and deep brain regions long implicated in affect in other mammals (e.g., rats). Technical development is still rapidly advancing, and many investigators have yet to avail themselves of the spatial and temporal resolution afforded by event-related fMRI. Accumulating evidence suggests that excessive averaging over either time or space can obscure meaningful fluctuations in subcortical activity. A parallel methodological challenge involves developing tools that can track ongoing changes in psychological phenomena, such as affect, with adequate temporal resolution, which could then be directly correlated with changes in brain activation. Hopefully in the future, not only will psychology inform neuroimaging, but neuroimaging will also reciprocally inform psychology. For instance, we have repeatedly observed that NAcc activation correlates with increases in positive arousal. Based on this association, one might predict that people will experience more positive arousal at any point in a given task when NAcc activation increases. This prediction has been confirmed in the case of the MID task, in which online affect probes have revealed that positive arousal increases most markedly during gain anticipation, rather than in response to gain outcomes (Knutson, Nielsen, et al., 2005). Additionally, linking brain activity to function may help drive predictions about how current brain activity could influence subsequent behavior. For instance, emerging evidence suggests that NAcc activation may promote gain-seeking financial behavior, as well as cooperation with friends and punishment of enemies. Of course, the difficult work of testing alternative functional accounts of brain activation must also proceed. In the case of affect, this endeavor will require designs that utilize subjectively compelling stimuli while controlling for sensorimotor demands, valence, arousal, and anticipation.

Another part of the coming revolution will be theoretical. Neuroimaging has revealed a beautiful and elegant confluence of results. The NAcc is activated not only by anticipation of nonsocial rewards such as money, food, or pleasant sensations but also by anticipation of social rewards, such as interacting with one's child or lover, rewarding a friend, or punishing an enemy. Thus neuroimaging findings may bring together previously disparate domains of inquiry and reveal surprising and useful connections between them. The current data suggest that social interaction is a multilayered process that powerfully involves affect. By implication, theories that attempt to describe and predict social interaction should strive to assign a central role for affect.

Although surprisingly coherent, the existing findings raise more questions than they answer. If a mechanism for gain anticipation responds to

social incentives and modulates subsequent social behavior, what other motivational circuits also play a role, how many are there, and how do they dynamically interact? At minimum, based both on brain stimulation data (Panksepp, 1998) and the statistical independence of self-reported positive arousal and negative arousal (Watson & Tellegen, 1985), one might postulate a loss anticipation system (Kuhnen & Knutson, 2005). Candidate regions might include those involved in the anticipation of pain, such as the periaqueductal gray, ventromedial hypothalamus, lateral amygdala, anterior insula, and anterior cingulate (Panksepp, 1998). Different neurochemicals might modulate activity of these distinct circuits (Depue & Collins, 1999; Knutson et al., 1998). Further, what kinds of control systems govern or modulate the activity of these deep motivational systems, how many are there, and how do they dynamically interact? For instance, our findings suggest that in the case of gain anticipation, the mPFC may modulate activity in the NAcc and enable people to correct their expectations of reward when violated or to keep probabilistic concerns in mind (Knutson, Taylor, et al., 2005). Finally, do social incentives recruit these systems differently than nonsocial incentives, and what additional circuitry promotes social interaction?

The goal of this overview is not to provide an exhaustive survey but, rather, to highlight a hypothesis that may prove useful to social neuroscientists. The hypothesis that NAcc activation indexes gain anticipation (and thereby generates an appetitive signal) provides a simple but powerful unifying framework both for consolidating diverse findings and also for generating predictions about future cognition and behavior. The hypothesis generalizes to both nonsocial and social rewards. It not only predicts where in the brain and when in time activation will occur but also how brain activation might influence subsequent behavior. At present, we know something, but not much. By applying neuroscience tools to the study of social behavior, we feel certain in predicting that we have much to gain.

ACKNOWLEDGMENTS

Brian Knutson was supported by a National Alliance on Schizophrenia and Depression Young Investigator Award and National Institute of Aging Seed Grant No. AG024957 during manuscript preparation. We thank Meghana Bhatt, Jeffrey Cooper, and Jeanne L. Tsai for helpful comments on previous drafts.

REFERENCES

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron, 50,* 507–517.

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11,* 231–239.

Aharon, I., Etcoff, N. L., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron, 32*, 537–551.

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G. H., et al. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6*, 196–202.

Arnow, B., Desmond, J. E., Banner, L. L., Glover, G. H., Solomon, A., Polan, M. L., et al. (2002). Brain activation and sexual arousal in healthy, heterosexual males. *Brain, 125*, 1–10.

Aron, A., Fisher, H. E., Mashek, D. J., Strong, G., Li, H. F., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology, 94*, 327–337.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390–1396.

Bandura, A. (1977). *Social learning theory*. Englewood Clifs, NJ: Prentice Hall.

Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *NeuroReport, 11*, 3829–3834.

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage, 21*, 1155–1166.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*, 122–142.

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience, 21*, 2793–2798.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309–369.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. *Games and Economic Behavior, 25*, 424–459.

Bishop, M. P., Elder, S. T., & Heath, R. G. (1963). Intracranial self-stimulation in man. *Science, 140*, 394–396.

Bjork, J. M., Knutson, B., Fong, G. W., Caggiano, D. M., Bennett, S. M., & Hommer, D. W. (2004). Incentive-elicited brain activation in adolescents: Similarities and differences from young adults. *Journal of Neuroscience, 24*, 1793–1802.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron, 30*, 619–639.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron, 17*, 875–887.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J. R., & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science, 304*, 1167–1170.

Craig, W. (1918). Appetites and aversions as constituents of instincts. *Biological Bulletin, 34*, 91–107.

Davidson, R., & Sutton, S. (1995). Affective neuroscience: The emergence of a discipline. *Current Opinions in Neurobiology, 5*(2), 217–224.

de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science, 305*, 1254–1258.

Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C., & Fiez, J. A. (2000). Tracking the hemodynamic response to reward and punishment in the striatum. *Journal of Neurophysiology, 84*, 3072–3077.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences, 22*, 491–517.

Dickinson, A. (1994). Instrumental conditioning. In N. J. Mackintosh (Ed.), *Animal learning and cognition* (pp. 45–79). New York: Academic Press.

Drevets, W. C., Gautier, C., Price, J. C., Kupfer, D. J., Kinahan, P. E., Grace, A. A., et al. (2001). Amphetamine-induced dopamine release in human ventral striatum correlates with euphoria. *Biological Psychiatry, 49*, 81–96.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience, 20*, 6159–6165.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., et al. (1994). fMRI of human visual cortex. *Nature, 369*, 525.

Falck, B., & Hillarp, N. A. (1959). On the cellular localization of catecholamines in the brain. *Acta Anatomica, 38*, 277–279.

Ferretti, A., Caulo, M., Del Gratta, C., Di Matteo, R., Merla, A., Montorsi, F., et al. (2005). Dynamics of male sexual arousal: Distinct components of brain activation revealed by fMRI. *NeuroImage, 26*, 1086–1096.

Gallagher, H., & Frith, C. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Science, 7*, 77–83.

Garavan, H., Pankiewicz, J., Bloom, A., Cho, J. K., Sperry, L., Ross, T. J., et al. (2000). Cue-induced cocaine craving: Neuroanatomical specificity for drug users and drug stimuli. *American Journal of Psychiatry, 157*, 1789–1798.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*, 10829–10837.

Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology, 108*, 233–242.

Hamann, S. B., Herman, R. A., Nolan, C. L., & Wallen, K. (2004). Men and women differ in amygdala response to visual sexual stimuli. *Nature Neuroscience, 7*, 411–416.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews, 31*, 6–41.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kampe, K. K., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Reward value of attractiveness and gaze. *Nature, 413*, 589.

Kampe, K. K., Frith, C. D., Dolan, R. J., & Frith, U. (2002). Correction. *Nature, 416*, 602.

Karama, S., Lecours, A. R., Leroux, J. M., Bourgouin, P., Beaudoin, G., Joubert, S.,

et al. (2002). Areas of brain activation in males and females during viewing of erotic film excerpts. *Human Brain Mapping, 16*, 1–13.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*, 78–83.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*, 165–182.

Knutson, B. (2004). Sweet revenge? *Science, 305*, 1246–1247.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience, 21*, RC159.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurobiology, 18*, 411–417.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related FMRI. *NeuroReport, 12*, 3683–3687.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related FMRI. *NeuroImage, 18*, 263–272.

Knutson, B., Nielsen, L., Larkin, G., & Carstensen, L. L. (2006). *Affect dynamics: Tracking trajectories through affect space.* Manuscript submitted for publication.

Knutson, B., Taylor, J., Kaufman, M. T., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience, 25*, 4806–4812.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage, 12*, 20–27.

Knutson, B., Wolkowitz, O. M., Cole, S. W., Chan, T., Moore, E. A., Johnson, R. C., et al. (1998). Selective alteration of personality and social behavior by serotonergic intervention. *American Journal of Psychiatry, 155*, 373–379.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., et al. (1998). Evidence for striatal dopamine release during a video game. *Nature, 393*(6682), 266–268.

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk-taking. *Neuron, 47*, 620–628.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral and behavioral reactions. *Psychophysiology, 30*, 261–273.

MacLean, P. D. (1990). *The triune brain in evolution.* New York: Plenum Press.

Mawlawi, O., Martinez, D., Slifstein, M., Broft, A., Chatterjee, R., Hwang, D., et al. (2001). Imaging human mesolimbic dopamine transmission with positron emission tomography: I. Accuracy and precision of D2 receptor parameter measurements in ventral striatum. *Journal of Cerebral Blood Flow and Metabolism, 21*, 1034–1057.

McBride, W. J., Murphy, J. M., & Ikemoto, S. (1999). Localization of brain reinforcement mechanisms: Intracranial self-administration and intracranial place-conditioning studies. *Behavioral Brain Research, 101*, 129–152.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional

imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the USA, 98*, 11832–11835.

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: The modern role of fMRI. *Neuroscientist, 10*, 260–268.

Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *NeuroImage, 28*, 175–184.

Mogenson, G. M., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology, 14*, 69–97.

Mouras, H., Stoleru, S., Bittoun, J., Glutron, D., Pelegrini-Issac, M., Paradis, A., et al. (2003). Brain processing of visual sexual stimuli in healthy men: A functional magnetic resonance imaging study. *NeuroImage, 20*, 855–869.

Nitschke, J. B., Nelson, E. E., Rusch, B. D., Fox, A. S., Oakes, T. R., & Davidson, R. J. (2004). Orbitofrontal cortex tracks positive mood in mothers viewing pictures of their newborn infants. *NeuroImage, 21*, 583–592.

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., & McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *Journal of Neurophysiology, 85*, 1315–1321.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Current Opinion in Neurobiology, 14*, 769–776.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron, 33*, 815–826.

O'Doherty, J. P., Winston, J. S., Critchley, H. D., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia, 41*, 147–155.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47*, 419–427.

Olds, M. E., & Fobes, J. L. (1981). The central basis of motivation: Intracranial self-stimulation studies. *Annual Review of Psychology, 32*, 523–574.

Panksepp, J. (1991). Affective neuroscience: A conceptual framework for the study of emotions. In K. Strongman (Ed.), *International reviews of studies in emotions* (Vol. 1, pp. 59–99). Chichester, UK: Wiley.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Panksepp, J., Knutson, B., & Burgdorf, J. (2002). The role of emotional brain systems in addictions: A neuro-evolutionary perspective. *Addiction, 97*, 459–469.

Pappata, S., Dehaene, S., Poline, J. B., Gregoire, M. C., Jobert, A., Delforge, J., et al. (2002). In vivo detection of striatal dopamine release during reward: A PET study with [(11)C]raclopride and a single dynamic scan approach. *NeuroImage, 16*, 1015–1027.

Park, K., Seo, J. J., Kang, H. K., Ryu, S. B., Kim, H. J., & Jeong, G. W. (2001). A new potential of blood oxygenation level dependent (BOLD) functional MRI for evaluating cerebral centers of penile erection. *International Journal of Impotence Research, 13*, 73–81.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford, UK: Oxford University Press.

Ramnani, N., Elliott, R., Athwal, B. S., & Passingham, R. E. (2004). Prediction error for free monetary reward in the human prefrontal cortex. *NeuroImage, 23*(3), 777–786.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A neural basis for social cooperation. *Neuron, 35*, 395–405.

Rolls, E. T., O'Doherty, J., Kringelbach, M. L., Francis, S., Bowtell, R., & McGlone, F. (2003). Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cerebral Cortex, 13*, 308–317.

Rotter, J. B. (1972). *Applications of a social learning theory of personality*. New York: Holt.

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

Semendeferi, K., Lu, A., Schenker, N., & Damasio, H. (2002). Humans and great apes share a large frontal cortex. *Nature Neuroscience, 5*, 272–276.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*, 653–662.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.

Volkow, N. D., Wang, G., Fowler, J. S., Logan, J., Gatley, S. J., Wong, C., et al. (1999). Reinforcing effects of psychostimulants in humans are associated with increases in brain dopamine and occupancy of D2 receptors. *Journal of Pharmacology and Experimental Therapeutics, 291*, 409–415.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–235.

Wightman, R. M., & Robinson, D. L. (2002). Transient changes in mesolimbic dopamine and their association with "reward." *Journal of Neurochemistry, 82*, 721–735.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45*, 459–467.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zald, D. H., Boileau, I., El-Dearedy, W., Gunn, R. N., McGlone, F., Dichter, G. S., et al. (2004). Dopamine transmission in the human striatum during monetary reward tasks. *Journal of Neuroscience, 24*, 4105–4112.

Nice work!