Why Losses Loom Larger than Gains: Modeling Neural Mechanisms of Cognitive-Affective Interaction

Abninder Litt (alitt@uwaterloo.ca)
School of Computer Science
University of Waterloo, ON, Canada N2L 3G1

Chris Eliasmith (celiasmith@uwaterloo.ca)
Departments of Philosophy and Systems Design Engineering
University of Waterloo, ON, Canada N2L 3G1

Paul Thagard (pthagard@uwaterloo.ca)
Department of Philosophy
University of Waterloo, ON, Canada N2L 3G1

Abstract

We present a biologically realistic spiking neural model that provides a broad-ranging mechanistic description of the human reward system. We introduce a novel conception of the role of affective arousal in stimulus valuation, and describe a dopamine-serotonin opponency in reward prediction error that influences both cognitive planning and emotional state. The model provides a neurological explanation of loss aversion in humans, and suggests particular mechanisms by which serotonin influences affective appraisal and risky behavior. Specific empirical predictions of the model include a correlation between amygdala serotonin receptor concentration and loss sensitivity, as well as specialized impairments resulting from several disconnection disorders. Our results provide a basis for further exploration of the neuroscientific foundations of economics, decision making and social cognition.

Keywords: computational neuroscience; dopamine; emotion; loss aversion; reward; serotonin.

Introduction

The reward system underlies the interactions between motivation, affect and cognition that give rise to decision making and goal-directed behavior. Understanding its structure and operation provides insight into addiction, preference and choice, and other complex environmental interactions. Mounting experimental data continue to elaborate the neural circuitry involved in the evaluation, prediction and behavioral effects of reinforcing stimuli (e.g., Montague & Berns, 2002). In humans and higher-order primates, much evidence indicates the involvement of such brain regions as the amygdala, orbitofrontal cortex, ventral striatum, anterior cingulate cortex, and dorsolateral prefrontal cortex in reward processing and related goal representation (as reviewed by McClure, York & Montague, 2004, and Schultz, 2000).

While identification of the key neural structures of the reward system has been greatly advanced by recent imaging and neurophysiological studies, the formulation of *mechanisms* by which these brain areas might interact has lagged behind this data accumulation. Theoretical work has focused on modeling restricted subsystems, such as dopamine-encoded reward prediction error (Schultz, 1998; Suri, 2002) and reward association reversal in orbitofrontal cortex (Deco & Rolls, 2005), rather than the large-scale synthesis required for moving beyond task-specific explanations to a more comprehensive neural theory of motivated behavior. The model we present is a first attempt at providing this kind of broad mechanistic description of the human reward system. It integrates current ideas about how different reward processing tasks are performed by the brain in a precise and quantitative manner, and provides an organizing basis for future experimental and theoretical study of reward-related behavior.

A key outcome of this integration is a novel and detailed neural basis for loss aversion. That people generally prefer avoiding losses to achieving equal gains was convincingly established by the landmark studies of Kahneman and Tversky (Kahneman & Tversky, 1979; Tversky & Kahneman, 1991). A psychological dominance of losses is thought to underlie many observed inconsistencies with traditional economic theory, including asymmetric price elasticity, downward-sloping labor supply and consumer choice anomalies (e.g., Camerer, 2000). Our model offers a biologically realistic neural explanation of a phenomenon studied almost exclusively at the behavioral level.

In brief, we describe a modulation of reward valuation by emotional arousal, influenced by stimulus saliency. This modulated signal feeds into interacting opponent systems for determining positive and negative errors in reward prediction. The consolidation of this information about valuation, saliency and relative surprise drives the planning of stimulus-appropriate behavior. Finally, information regarding behavioral relevance and prediction error feeds back to modulate arousal level itself. Simulation results suggest neural explanations for a wide range of phenomena, such as loss aversion and the behavioral influence of serotonin. Above all, our model demonstrates that diverse data regarding the neural substrates of reward processing can be successfully integrated, explaining activities and behaviors that are themselves combinations of the observed and theorized functions of these subsystems.
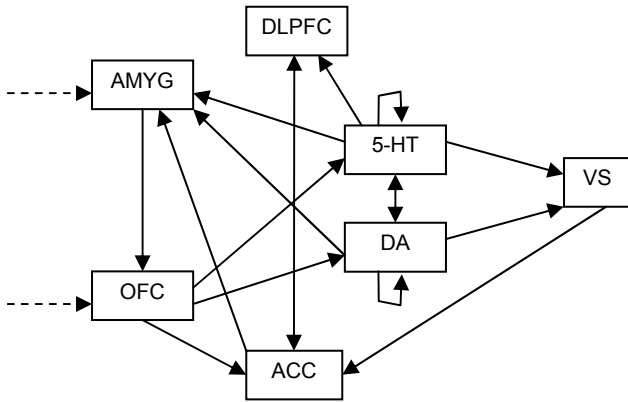
Model Architecture

Figure 1: Basic connectivity framework. Dotted arrows represent external inputs to the model. Abbreviations: 5-HT, dorsal raphe serotonergic neurons; ACC, anterior cingulate cortex; AMYG, amygdala; DA, midbrain dopaminergic neurons; DLPFC, dorsolateral prefrontal cortex; OFC, orbitofrontal cortex; VS, ventral striatum.

Figure 1 outlines the overall structure of our model. It consists of 7600 spiking, leaky integrate-and-fire (LIF) neurons in 7 distinct populations. Specifically, we use 800-1200 neurons for simulating each of the amygdala, orbitofrontal cortex, ventral striatum, anterior cingulate cortex, and dorsolateral prefrontal cortex. The activities of midbrain dopaminergic neurons and the dorsal raphe nucleus of the brainstem are modeled with 1200-neuron ensembles, each with several discrete subpopulations. We employ the Neural Engineering Framework (NEF) of Eliasmith and Anderson (2003) implemented through the MATLAB-based neural simulation software *NESim*.

We will now describe the basic relations and activities modeled for each brain area, and how they relate to the generation of complex reward-related behaviors. A complete description of our specific model equations and neural parameters, sufficient for thorough analysis and replication, is provided by Litt, Eliasmith and Thagard (2006) and omitted here for reasons of space.

Arousal Modulation of Reward Valuation

A wide range of studies have implicated orbitofrontal cortex in the evaluation of the rewarding nature of stimuli (reviewed by Rolls, 2000). With myriad connections to primary and secondary sensory areas, it is well-placed to provide an integrative valuation of the rewarding or punishing nature of external inputs. This valuation is influenced by context (Schultz, 2000) as well as interactions with neural structures involved in emotional processing (Bechara, Damasio & Damasio, 2000). We propose that a primary modulator of orbitofrontal reward valuation, which may underlie both emotional and context-sensitive influences, is the state of affective *arousal* of the individual, which we model via activity in the amygdala.

The conventional association of the amygdala with the processing of primarily aversive stimuli is being challenged by a range of studies which indicate that it is the degree to which a stimulus is arousing, rather than its valence, that is related to amygdala activation (McClure, York & Montague, 2004). Classic results can be reinterpreted as showing that aversive stimuli are generally more arousing than rewarding ones, perhaps because of increased behavioral saliency. Negative feedback often induces an individual to modify current behavior, demanding more complex cognitive processing, such as plan formation, which is not required in the case of positive feedback. This may partially underlie the valence asymmetry in amygdala activation. The finding of Adolphs and colleagues (2005) that the impairment of fear recognition caused by amygdala damage is due to abnormal negligence of the eye region of faces also indicates an attentional role for the amygdala.

In this spirit, we look to theoretical models of attention in developing our proposed interaction between the amygdala and orbitofrontal reward valuation. Evidence from vision research points to a *multiplicative scaling* effect of attention on the observed saliency of visual inputs (e.g., Treue, 2001). We thus propose a multiplicative modulation by arousal of stimulus valuation in orbitofrontal cortex. Given V and A as the neural representations of primary orbitofrontal valuation and amygdala-encoded emotional arousal, respectively, we model the modulated valuation output $V^*(t) = A(t) \cdot V(t)$. Thus, increased levels of arousal amplify stimulus valuation, while lower arousal levels lead to signal attenuation.

Opponency in Reward Prediction Error

An extensive body of data implicates midbrain dopamine neurons in reporting a *reward prediction error* that reflects the discrepancy between actual and expected results (e.g., Schultz, 2000; Suri, 2002). The temporal difference (TD) model (Sutton & Barto, 1998) is the dominant theoretical approach for this computation, due to simplicity and good correspondence to observed neural activity. TD computes reward prediction error (E) based on the difference between the latest reward valuation and a weighted sum of all previous rewards (P). Using our arousal-modulated signal V^* as the input regarding current stimulus valuation, this leads to the modeled equations $E(t) = V^*(t) - P(t - 1)$ and $P(t) = P(t - 1) + \alpha \cdot E(t)$, where α is a learning rate constant between 0 and 1. This is typically modeled by increasing dopamine activity with positive prediction errors (that is, getting more than expected) and firing rate depression for negative errors (Schultz, 2000). While this has indeed been observed within certain ranges, the capacity of dopamine to work alone in this manner has recently been questioned. Daw, Kakade and Dayan (2002) argue that low baseline firing rates make a dopamine-only scheme unsuitable for computing highly negative prediction errors. Indeed, strongly aversive stimuli *increase* dopamine firing rates in a manner inconsistent with the standard picture (Horvitz, 2000). Recent experiments show that, while dopamine seems to encode positive prediction errors, this is not true for significantly negative errors (Bayer & Glimcher, 2005).

Daw and colleagues propose serotonin innervation by the dorsal raphe nucleus as working in interacting opponency with midbrain dopamine, with each system reacting primarily to aversive or appetitive stimuli, respectively. Among the diverse effects of serotonin in the brain is

consistent evidence of responsiveness in aversive situations (Deakin, 1983) and relatedly in behavioral inhibition, such as withholding response when a shock is the expected result (Soubrié, 1986). These roles seem naturally opposed to those commonly ascribed to dopamine, and Daw, Kakade and Dayan (2002) further outline direct physiological evidence for opponency between these two systems.

Our modeled opponency simplifies that proposed by Daw and colleagues. While this does not allow explanation of some more complex findings (Horvitz, 2000), it still divides positive and negative prediction error encoding, in line with recent neurological results. Litt, Eliasmith and Thagard (2006) present our specific opponent computations in the midbrain and raphe. We consolidate these encodings in the ventral striatum, which is also implicated in reward prediction error processing (York & Montague, 2004).

A key advantage of opponent systems for positive and negative reward prediction error is that we can distinctly calibrate outputs from these systems to other brain areas. Because prediction error is in effect a measurement of surprise, we hypothesize that one target of such outputs is the amygdala, which we describe as encoding arousal and stimulus saliency. We thus model feedback to the amygdala of the interacting opponent reward prediction error signals: $A(t) = A_0(t) + \beta \cdot DA(t) + \gamma \cdot 5\text{-}HT(t)$, A_0 being a base arousal level determined by external factors unrelated to reward. Importantly, the effects on arousal of positive and negative surprises can be *asymmetric* (i.e., $\beta \neq \gamma$). We explore the implications of this asymmetry in our later discussion of how our model explains loss-averse behavior in humans.

Conflict, Motivation and Behavior

Finally, we model the influence on behavioral planning of the reward processing activities discussed so far. The dorsolateral prefrontal cortex seems crucial for the planning, representation and selection of goal-directed behaviors (Hikosaka & Watanabe, 2000; Owen, 1997). The full extent of these tasks undoubtedly involves not only reward feedback but also information about body state, environmental constraints and myriad other factors. We restrict ourselves here to reward-related mechanisms of influence on the development of appropriate action plans.

In this regard, research has focused on the role of the anterior cingulate cortex in emotional consolidation, detecting conflicts between current behavior and desired results and interfacing with dorsolateral prefrontal (Bush, Luu & Posner 2000). In particular, neural activity dubbed error-related negativity (ERN) is vital for conflict detection and mediation, triggering relevant modifications to behavior planning in other cognitive areas (Brown & Braver, 2005). Our model proceeds along similar lines, and also proposes feedbacks to other stages of reward processing that help explain psychological and neuroimaging results. We have described elsewhere the precise mathematical nature of these modeled relations (Litt, Eliasmith & Thagard, 2006).

The consolidation implemented in the anterior cingulate merges ideas from the two dominant theories of neural emotional processing: the approach-withdrawal model (reviewed by Sutton, 2002) and the valence-arousal model (Heller, 1993). The amygdala-modulated reward valuation signal from orbitofrontal cortex is analyzed in the anterior cingulate for an appropriate behavioral response: approach, if the current input is valued positively; or withdrawal, if it is considered aversive. We then modulate this behavioral valence by reward-related surprise—that is, prediction error. If a given behavior (approach or withdrawal) occurs concurrently with a positive prediction error, the individual is receiving encouraging feedback with regard to that behavior, which should therefore be strengthened (or at least maintained). Conversely, negative prediction error indicates *conflict*: the behavior is not producing the expected result. We model a consequent weakening of the behavior via cingulate activity attenuation (Litt, Eliasmith & Thagard, 2006), as a new plan of action may need to be formulated.

Presumably, dorsolateral prefrontal cortex would use this data (as well as inputs from other brain areas) in planning and producing goal-directed behavior. While we do not model this process, we propose another interaction between this region and the cingulate. We propose a raphe-dorsolateral-cingulate-amygdala pathway by which negative prediction errors can further influence arousal. Physiological support for this pathway is provided by Wilson and Molliver (1991) and Bush and colleagues (2000). As explained earlier, negative surprise (and corresponding raphe serotonergic activity) indicates conflict and the possible need for behavioral modification. We postulate an arousal increase as a consequence of this behavioral saliency, due to the additional cognitive resources such a situation would demand. Using the described pathway, the intermediate steps produce increased activity in the cingulate that may relate to the conflict-related ERN activity. Labeling the output of this pathway C, the cost of conflict and potential behavior modification, we obtain a final description of our model's modulation of emotional arousal level: $A(t) = A_0(t) + \beta \cdot DA(t) + \gamma \cdot 5\text{-}HT(t) + C(t)$.

A Neural Basis for Loss Aversion

While the psychology of loss aversion has been studied extensively, little work has focused on identifying its neural foundations. Loss aversion is principally a stimulus valuation phenomenon (an overvaluation of negatively construed stimuli). A neurological explanation should thus arise from mechanisms involved in the reward system. We illustrate how our model accounts for loss-averse behavior, specifically in terms of the asymmetric orbitofrontal cortex valuations observed in our experimental simulation results.

We outline mechanisms by which arousal is enhanced disproportionately by negative rewards as opposed to positive equivalents. This leads to overvaluation of negative stimuli, by our multiplicative arousal modulation procedure. Thus, one's attention is captured more by negative events, which in turn causes them to loom larger than gains. One way in which losses are made additionally arousing is our calibration asymmetry between dopamine-encoded positive prediction error and serotonin-encoded negative error (i.e., serotonin has stronger modulatory effects on the amygdala than dopamine). This leads to a greater impact on arousal of negative versus positive outcomes, as required. Such disparity might have an evolutionary basis: negatively appraised events may often jeopardize survival and further

reproduction, while the same vital saliency does not commonly accompany boons (Nettle, 2005; Seligman, 2002). Simple realization of this uneven calibration is a favorable outcome of opponent prediction error encoding modeled initially for reasons of biological plausibility.

The other way losses additionally bias arousal is via the conflict cost C fed back to the amygdala from the cingulate and dorsolateral prefrontal areas. It is realistic to expect stimulus saliency, expressed through arousal level, to be enhanced by the necessity of modifying current behavior in response to the stimulus. Cognitive resources are needed to plan and select new actions, and enacting any changes may incur risks and energy costs. In general, it is less demanding and risky for an organism to continue its current course than to formulate and implement a new plan of action. We model this through our signaling of behavioral modification needs by *negative* prediction error input to the cingulate, which consequently detects conflicts between expected results (i.e., current goals) and reality that call for action adjustments. The cost is thus charged with perceived losses but not gains, so its feedback to the amygdala further augments the disparity between the saliency of losses and gains.

losses being exaggerated over gains. Our chosen model parameters lead to arousal signal amplitudes approximately 200% greater for losses than equal gains. Also note reward overvaluation caused by *both* losses and gains (which do so to different degrees) is transient, subsiding over time to restore more accurate valuations. This supports findings that loss aversion is often a knee-jerk affective reaction: decisions made with increased scrutiny and contemplation are often closer to rational agent behavior (Kahneman & Tversky, 1979). The nature of loss aversion is thus well replicated by our model, which proposes the first detailed, biologically realistic neural basis for the phenomenon.

Effects of Lesions to the Serotonin System

It is useful to examine the specific effects of introducing interactions with serotonergic neurons in the dorsal raphe, as this is a major divergence from typical conceptions of the reward system present in our model. We do so by ablating the area and observing how the operation of the model is affected. The outcomes of this lesioning agree well with the findings of experimental studies, and considerations of the overall model provide possible neurological mechanisms for the associated effects.
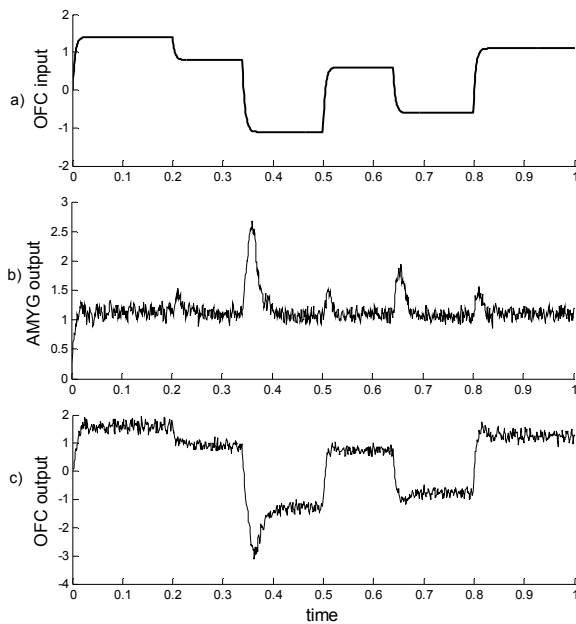

Figure 2: Asymmetric evaluation of gains and losses. Decoded output from spiking neuron populations (see Eliasmith & Anderson, 2003). **a)** The external input signal to orbitofrontal cortex consists of positive and negative valuation changes of varying magnitude. **b)** Arousal modulated by prediction error and the behavioral saliency of stimuli. Losses induce markedly greater arousal increases than gains. **c)** The outcome of the unevenness displayed in b). Stimulus valuation reductions (losses) are amplified disproportionately when compared to equivalent gains.

Loss-sensitive reward valuation by our model is vivid in the simulation of Figure 2. Asymmetric arousal increase caused by positive and negative valuation changes leads to
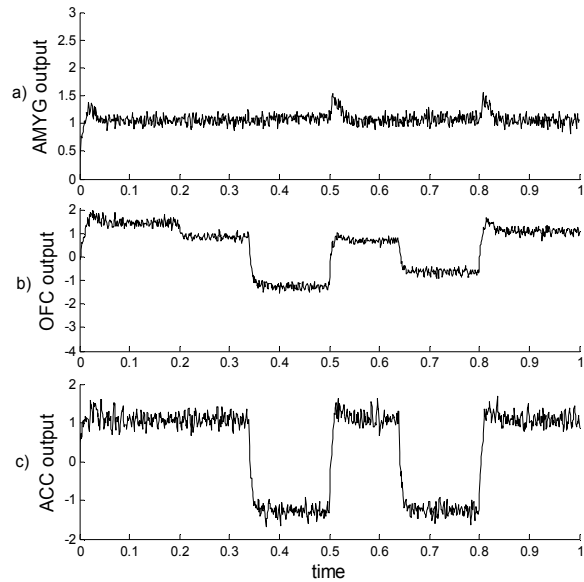

Figure 3: Serotonin lesioning. Data generated for the same external input signal used in Figure 2, with the dorsal raphe now removed. **a)** Arousal is no longer amplified by unexpected negative stimuli, as that measurement and reporting system has been lesioned (cf. Fig. 2b). **b)** Loss-averse reward valuation ceases, as the necessary upsurges in the arousal modulator no longer occur (cf. Fig. 2c). **c)** Unaffected is whether approach or withdrawal is favored by the cingulate based on current reinforcer valence, but action inhibition subsequent to receiving negative feedback about current behavior has been eliminated.

Figure 3 illustrates the results of a simulation run with the dorsal raphe nucleus ablated. Figures 3a and 3b show that

negatively appraised events no longer increase arousal and hence the corresponding stimulus valuation. Loss aversion disappears, as losses no longer provide any affective "jolt" to the individual. The impact of this is visible in the activity of the anterior cingulate (Fig. 3c). Response inhibition (i.e., signal diminishment towards 0) in the aftermath of a negative shift in reward disappears, as such losses are no longer accompanied by the serotonin release that previously occurred in the model and is present in real brains (Deakin, 1983; Rogers et al., 2003). This result agrees with animal lesion studies showing similar response inhibition disturbance (Harrison, Everitt & Robbins, 1999; Soubrié, 1986) as well as work with human subjects (Tanaka et al., 2004). In addition, the behavioral modification cost that was formerly fed back to increase arousal is no longer computed, as it depended primarily on the serotonin-encoded negative reward prediction error signal to detect situations that called for new plans of action. This mirrors the decrease in behavior-switching penalty in lesioned rats described by Al-Ruwaitea and colleagues (1999). The end result of these changes is that behavior is less cautious, which follows from negative feedback being essentially disregarded. Risky behavior is exactly that which ignores potential negative consequences of actions, so by damaging the system that computes and represents these penalties it is unsurprising that the result is an increase in such behavior.

Overall, lesioning the raphe serotonin system has the interlocking effects of damaging negative reward appraisal, eliminating the effect of behavior modification costs on arousal and reward valuation, and making behavior more risk-prone and heedless of penalty. By reproducing these experimental findings in our model, we are able to provide insight into their neurological foundations. In particular, the primary function we assign to serotonin in our model—computation and representation of negative reward prediction error—seems to underlie all of the described effects that accompany dorsal raphe lesioning. In terms of both direct affective impact and consequences to further behavioral planning, the encoding of loss-related information by serotonergic neurons seems vital to the normal operation of reward processing and executive control systems. In may be enlightening in future work to also explore complementary disruptions of the dopamine system, comparing model simulation results to behavioral observations regarding reinforcement through rewards and punishers (e.g., Frank, Seeberger & O'Reilly, 2004).

Conclusion

While there remains much to explore within smaller subsets of the reward system, explaining complex psychological phenomena requires a wider synthesis of the available experimental data. Outlining computations *between* relevant brain areas as well as those primarily performed locally represents a move towards larger-scale models. The specific modes of interaction we have implemented prove useful in providing wide-ranging, explanatorily fruitful neural bases for behavior motivated by affective stimulus valuation.

Besides providing explanations for existing data, our model also leads to specific experimental predictions regarding reward processing. Based on our connectivity structure (Fig. 1) we expect a positive correlation between serotonin receptor concentration in the amygdala and observed sensitivity to losses, as a result of our hypothesized arousal modulation of stimulus valuation. In addition, damaged innervation between midbrain dopamine neurons and the raphe serotonin system is predicted to increase the extent to which *both* positive and negative rewards are initially overvalued, due to the mutual interactivity between these systems we have modeled. This may be tested behaviorally or verified through imaging data showing increased amygdala and orbitofrontal activation during stimulus valuation tasks. Our modeling of indirect connectivity of orbitofrontal cortex to the amygdala through the raphe and midbrain subsystems is a further source of predictions. We expect orbitofrontal damage to produce degraded context-sensitive affective arousal, reflected neurologically via disturbed amygdala activation in situations of rewarding or punishing environmental feedback. The specific cause of this disturbance is predicted to be an impairment of basic stimulus valuation, rather than such appraisal simply not influencing emotional saliency. Finally, while we have not modeled in detail plan formation in dorsolateral prefrontal cortex, connectivity through this region employed by our neural architecture can inform testable hypotheses. One such prediction is that disturbed innervation from raphe serotonin to dorsolateral prefrontal cortex will cause a particular impairment in *acting* upon negative environmental feedback (e.g., planning and implementing appropriate behavioral modifications), while sparing capabilities to recognize and appraise such feedback. More general plan formation and execution abilities should also be preserved, as our model predicts these impairments entail *direct* damage to dorsolateral prefrontal areas, not simply degraded serotonin innervation. The ability to provide this wide range of predictions stems from the large-scale nature of our model, and illustrates the benefits of taking an extensive-scope approach.

Future applications of this broad approach could include further exploration of the neuropsychological foundations of economics and decision making. While loss aversion is itself an important and complex phenomenon, it seems also to underlie even more elaborate behaviors. The sunk-cost fallacy (Arkes & Ayton, 1999), framing effects (Tversky & Kahneman, 1981) and the prisoner's dilemma (Axelrod & Hamilton, 1981) are just a few of the important economic phenomena that stand to be enlightened by a neuroscientific approach. In addition, a characterization of risk as feeling (Loewenstein et al., 2001) would enable the exploration of risk calculation, appraisal and mitigation mechanisms under the affective valuation-modulation paradigm we have developed. Large-scale modeling of the reward system, such as that which we have described here, opens many areas at the boundaries of emotion, cognition, and social interaction to in-depth investigation at the neurological level.

Acknowledgments

References

Adolphs, R., et al. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*, 68-72.

Al-Ruwaitea, A. S. et al. (1999). Effect of central 5-hydroxytryptamine depletion on changeover behaviour in concurrent schedules of reinforcement. *Psychopharmacology (Berl.), 144*, 264-271.

Arkes, H. R. & Ayton, P. (1999). The sunk cost and Concorde effects: are humans less rational than lower animals? *Psychological Bulletin, 125*, 591-600.

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

Bayer, H. M. & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron, 47*, 129-141.

Bechara, A., Damasio, H. & Damasio, A. R. (2000). Emotion, decision-making, and the orbitofrontal cortex. *Cerebral Cortex, 10*, 284-294.

Brown, J. W. & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*, 1118-1121.

Bush, G., Luu, P. & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*, 215-222.

Camerer, C. F. (2000). Prospect theory in the wild: evidence from the field. In D. Kahneman & A. Tversky (Eds.), *Choices, Values, and Frames*. New York: Cambridge University Press.

Daw, N. D., Kakade, S. & Dayan, P. (2002). Opponent interactions between serotonin and dopamine. *Neural Networks, 15*, 603-616.

Deakin, J. F. W. (1983). Roles of brain serotonergic neurons in escape, avoidance and other behaviors. *Journal of Psychopharmacology, 43*, 563-577.

Deco, G. & Rolls, E. T. (2005). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cerebral Cortex, 15*, 15-30.

Eliasmith, C. & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science, 306*, 1940-1943.

Harrison, A. A., Everitt, B. J. & Robbins, T. W. (1999). Central serotonin depletion impairs both the acquisition and performance of a symmetrically reinforced go/no-go conditional visual discrimination. *Behavioural Brain Research, 100*, 99-112.

Heller, W. (1993). Neuropsychological mechanisms of individual differences in emotion, personality, and arousal. *Neuropsychology, 7*, 476-489.

Hikosaka, K. & Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cerebral Cortex, 10*, 263-271.

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience, 96*, 651-656.

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

Litt, A., Eliasmith, C. & Thagard, P. (2006). *A large-scale neural model of human reward processing*. Manuscript submitted for publication.

Loewenstein, G. F., Weber, E. U., Hsee, C. K. & Welch, E. (2001). Risk as feelings. *Psychological Bulletin, 127*, 267-286.

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

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

Nettle, D. (2005). *Happiness*. Oxford: Oxford University Press.

Owen, A. M. (1997). Cognitive planning in humans: neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology, 53*, 431-450.

Rogers, R. D. et al. (2003). Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. *Neuropsychopharmacology, 28*, 153-162.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex, 10*, 284-294.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1-27.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience, 1*, 199-207.

Seligman, M. E. P. (2002). *Authentic Happiness*. New York: Free Press.

Soubrié, P. (1986). Reconciling the role of central serotonin neurons in human and animal behavior. *Behavioral and Brain Sciences, 9*, 319-335.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Networks, 15*, 523-533.

Sutton, S. K. (2002). Incentive and threat reactivity: relations with anterior cortical activity. In D. Cervone & W. Mischel (Eds.), *Advances in personality science*. New York: Guilford Press.

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

Tanaka, S. C. et al. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience, 7*, 887-893.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24*, 295-300.

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

Tversky, A. & Kahneman, D. (1991). Loss aversion in riskless choice: a reference dependent model. *Quarterly Journal of Economics, 106*, 1039-1061.

Wilson, M. A. & Molliver, M. E. (1991). The organization of serotonergic projections to cerebral cortex in primates: retrograde transport studies. *Neuroscience, 44*, 555-570.