When Things Get Worse before they Get Better: Regulatory Fit and Average-Reward Learning in a Dynamic Decision-Making Environment

A. Ross Otto, Arthur B. Markman, and Bradley C. Love
Department of Psychology, University of Texas, Austin, TX 78712 USA
[rotto@mail.utexas.edu, markman@psy.utexas.edu, love@psy.utexas.edu]

Todd M. Gureckis
Department of Psychology, New York University, 6 Washington Place, New York, NY 10003
[todd.gureckis@nyu.edu]

Abstract

This work explores the influence of motivation on choice behavior in a dynamic decision-making environment, where the payoffs from each choice depend on one's recent choice history. Previous research reveals increased levels of exploratory choice among participants in a regulatory fit. The present study placed promotion and prevention-focused participants in a dynamic environment for which optimal performance requires that participants sustain a single choice strategy in the face of temporary payoff decreases. These participants either gained or lost points with each choice. Our behavioral results and model-based analysis, using the average-reward reinforcement learning framework, revealed differential levels of reactivity to local changes in payoffs—specifically, participants in a regulatory fit were less reactive to local perturbations in payoffs than participants in a regulatory mismatch and performed more optimally as a result.

Keywords: Decision making; motivation; reinforcement learning

Introduction

Motivation is essential to action (e.g., Carver & Scheier, 1998; Yerkes & Dodson, 1908). Social psychology makes the distinction between two general motivational orientations (or *regulatory foci*), a *promotion focus* and a *prevention focus*, which accentuate potential gains and losses in the environment, respectively (Higgins, 1997). Recent research reveals that an interaction occurs between one's regulatory focus and the reward structure of the task being performed, affecting peoples' use of flexible strategies in a number of tasks. In one study (Worthy, Maddox, & Markman, 2007) utilizing a two-armed bandit task for which optimal choice behavior required exploratory choices as opposed to exploitative choices (c.f. Daw et al., 2006), participants attempting to *earn* a prize (inducing a promotion focus) exhibited more optimal choice behavior when the task environment had a gains reward structure (i.e., participants were maximizing *gains* of points) than when the task environment had a losses reward structure (i.e. participants were minimizing *loss* of points). Likewise, participants attempting to avoid losing a prize (a prevention focus) performed more optimally when the task involved a losses reward structure than when the task involved a gains reward structure.

In *n*-armed bandit tasks where which the decision-maker learns to maximize his or her payoffs by making choices and experiencing the consequences of those choices (e.g. Daw et al., 2006; Bechara, A.R. Damasio, H. Damasio, & Anderson, 1994), optimal performance depends on balancing the demands of gathering and exploiting information about choice payoffs. Worthy et al.'s (2007) study demonstrated that participants in a *regulatory fit,* for whom their situational regulatory focus matches the reward structure of the task environment, exhibit more exploratory choice strategies than do participants in a *regulatory mismatch*, who exhibit more exploitative choice strategies. Through continued exploration of choices—with the consequence of occasionally taking decreases in payoffs—participants in a regulatory fit display behavior that is adaptive for the overall long-term pursuit of rewards. Further, their model-based analyses suggested that participants in a regulatory fit place less weight on outcomes from recent choices than do participants in a regulatory mismatch. While this class of tasks is well suited for investigating exploratory versus exploitative choice behavior, we seek to understand how motivational factors affect the degree to which recent changes in payoffs drive choice behavior.

In this report, we examine the effects of regulatory fit in a two-option, repeated-choice decision making task in which payoff-maximizing, long-term optimal behavior requires that participants persevere with one choice strategy, sustaining temporary decreases in payoffs in order to maximize long-term gain. Our experiment placed participants in a version of the "rising optimum" task, previously used to investigate temporal-difference accounts of learning (Eagelman, Person, & Berns, 1998; Montague & Berns, 2002) and the problem of temporal credit assignment in human sequential decision-making (Bogacz, McClure, Li, Cohen, & Montague, 2007). Unlike other bandit tasks in which payoff contingencies remain invariant to participants' behavior (e.g., Bechara et al., 1994; Daw et al., 2006; Worthy et al., 2007; Yechaim & Busemeyer; 2005), in this task, the *state* of the task environment changes as a function of a participant's recent choices, which in turn governs the payoffs associated with each action.

Consider the two payoff curves depicted in Figure 1, which correspond to the possible payoffs for two choices A and B in Egelman et al.'s (1998) "rising optimum" task. The payoff received from a choice depends on the proportion of A choices made over the last 20 trials, represented by the

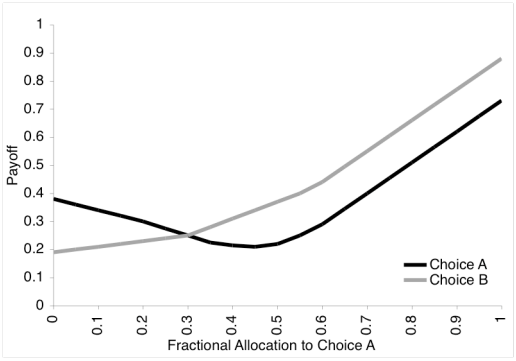

Figure 1: Payoff functions for two choices as a function of response allocation. (See text for details).

horizontal axis. For example, if the participant makes only B choices for 20 trials in a row—effectively making their fractional allocation to the A choice 0—the payoffs from choices A and B would be 0.38 and 0.19 respectively. If the participant makes one A choice at this point, his or her response allocation would change to 0.05, as only 1 out of 20 of the last trials were A choices. Consequently, the payoffs for choice A and B would be .36 and 0.2. Thus, the payoffs associated with the choices fluctuate with the past choice behavior of the participant. In this task, optimal long-term choice behavior requires consistent A choices every trial, as the global optimum is located where the participant's fractional choice allocation to choice A is 1 (Montague & Berns, 2002).

Prior research utilizing the rising optimum task reveals that participants easily become "stuck" in a local cycle around the crossing point of the curves where the fractional allocation to choice A is approximately 0.3 (Bogacz et al., 2007; Montague & Berns, 2002). To illustrate, consider a participant who makes repeated A choices until they find themselves at the crossing point of the two curves (Figure 1). As they continue to make A choices, exceeding an allocation of 0.3, the immediate payoff from choice A will decrease, with greater immediate payoffs resulting from choice B. Should they elect to make B choices at this point, rewards for that option will decrease until the fractional allocation falls below 0.3 whereupon choice A will yield higher immediate payoffs. This globally suboptimal response strategy—akin to matching behavior by humans described by Herrnstein (1990)—is predicted by simple temporal-difference (TD) learning models of reinforcement learning (Montague & Berns, 2002; Sutton & Barto, 1998). An optimal strategy of consistent A choice requires that the participant persist in the face of the local decrease in payoffs as they depart the "matching" crossing point and move towards the global optimum of the A payoff curve. In the absence of experimental manipulations, both Montague and Berns (2002) and Egelman et al. (1998) grouped participants by their choice strategies (i.e., those who stayed near the crossing point, and those who were able reach near-optimal allocations), finding that a substantial number of participants exhibited choice behavior not anticipated by standard TD-learning models.

A number of studies have explored factors that shape participants' choice allocations both in the rising optimum task and other similar dynamic environments. Bogacz et al. (2007) demonstrated how optimal choice performance depends on the amount of time that elapses between choices (i.e., inter-choice interval) using an eligibility trace model. A question then, comes to bear in light of prior research: how can motivational factors influence humans' pursuit of overall long-term rewards in the face of local reward decreases, consequently driving them toward or away from payoff-maximizing choice in the rising optimum task?

The present work extends previous research in two ways. First, we demonstrate that situational regulatory fit (and mismatch) affect the degree to which participants are able to sustain temporary decreases in payoffs in order to maximize long-term payoffs. Second, in the framework of reinforcement learning (RL), we provide a model-based analysis of choice behavior using a variant of the TD-learning algorithm (Sutton & Barto, 1998) known as average-reward learning, elucidating our hypothesized differences about participants' reactivity to local changes in payoffs. In short, we hypothesize that the interaction between one's motivational state and the reward structure of the environment will influence individuals' ability to sustain globally advantageous choices in the face of local perturbations, such that decision-makers in a regulatory fit will exhibit more optimal, payoff-maximizing response allocations than decision-makers in a regulatory mismatch.

Experiment 1

We placed participants in a variant of the Rising Optimum task, whose payoff schedule (under the gains reward structure) is depicted in Figure 1. Participants in the gains condition started with 0 points and gained between 0 and 1 points with each choice, while participants in the losses condition started with 0 points and lost between 0 and -1 points with each choice. The bonus criteria was positioned such that participants would need to earn at least 75% of the total possible points at the end of the experiment—which required that participants persevere in the face of local reward decreases as they made repeated A choices. Consequently, participants whose choice allocations remained near the "matching" equilibrium would not achieve the bonus criterion.

Participants in a promotion focus were told that they would receive an entry into a drawing for a 1 in 10 chance at winning $50 if they achieved the bonus criterion. Participants in a prevention focus were given an entry into the drawing and told that they had to achieve the bonus criterion to avoid losing the entry. As in previous research (e.g. Shaw & Higgins, 1997), this manipulation was designed so that participants in the promotion and prevention focus conditions were effectively in the same objective situation.

In light of previous work revealing heightened exploratory choice in regulatory fit (Worthy et al., 2007),

we hypothesized that participants in a regulatory fit (a promotion focus with a gains reward structure or a prevention focus with a losses reward structure) will sustain globally advantageous choice strategies, exhibiting less reactivity to local changes in payoffs. In contrast, we hypothesized that participants in a regulatory mismatch (a prevention focus with a gains reward structure or a promotion focus with a losses reward structure) would exhibit more reactivity to local changes in payoffs and thus exhibit less globally optimal response allocations. Table 1 provides another description of our factorial design and hypotheses.

Table 1: Overview of regulatory focus manipulation.

Regulatory Focus	*Reward Structure*	
	Gains	Losses
Promotion	**Fit** (decreased reactivity)	**Mismatch** (increased reactivity)
Prevention	**Mismatch** (increased reactivity)	**Fit** (decreased reactivity)

Method

Participants Forty undergraduates from the University of Texas community participated in the experiment for course credit. They were also given the opportunity to win an entry into a drawing for $50 cash, and were told that no more than 10 participants would be included in each drawing. The two between-subjects independent variables were the situational regulatory focus (promotion and prevention) and the reward structure of the task (gains and losses).

Materials The experiment stimuli and instructions were displayed on 17-inch LCD monitors. At the start of the experiment, participants were informed that they would either earn (promotion condition) or keep (prevention condition) a entry into the drawing if they met a bonus criterion. Participants were instructed to make repeated choices with the goal of maximizing overall, long-term gains of points (gains condition) or minimizing overall long-term losses of points (losses condition).

Procedure At the start of the experiment, each participant's response history was randomized such that the mean starting allocation of A choices was 0.5 across all participants. Each trial, participants were presented with two buttons labeled "Choice A" and "Choice B". The mapping of response buttons to choices was counterbalanced across participants. The task interface under the gains condition is shown in Figure 2. Using the mouse, participants clicked one of the buttons to indicate their choice, and white payoff bar grew (or fell, in the losses condition) vertically to indicate the amount of points gained (or lost, in the losses condition) on that trial. There was no time limit for making choices.

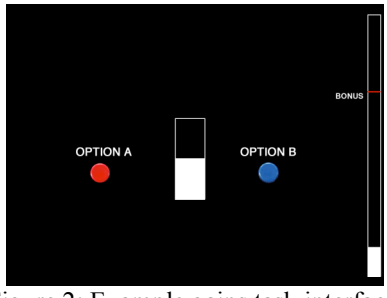

Figure 2: Example gains task interface.

The payoff each trial was a function of the relative fraction of the number of A choices made by the participant over the last 20 trials. Specifically, the payoff for each option, in the gains condition, with respect to relative fraction of A choices, is depicted in Figure 1. Gains payoffs were all between 0 and 1. Payoffs in the losses condition were calculated by subtracting 1 from the gains payoffs, resulting in all negative payoff values. Cumulative gains (or losses) were displayed on the side of the screen, as a bar that grew (or shrunk, in the losses condition) in relation to the bonus criterion. This bonus criterion was determined by calculating the average cumulative payoffs after 250 trials with an "A" choice allocation of 0.75. This criterion was equated across the gains and losses conditions.

After 250 trials, participants were given feedback on whether they had met the bonus criterion or not. If they met the bonus criterion, participants in the promotion focus condition were given a ticket and told to enter it in the drawing, and participants in the prevention focus condition were informed that they could keep their ticket and enter it in the drawing.

Results

Performance Measures

As a measure of response optimality, we analyzed the proportion of trials for which participants made optimal "A" choices. A 2 (regulatory focus) x 2 (reward structure) ANOVA conducted on overall proportion of A choices collapsed over the course of the experiment revealed a significant interaction ($F(1,38)=32.48$, $p<.001$) and no significant main effects. Among participants in the gains reward structure, participants in a promotion focus ($M=0.591$, $SD=0.05$) made significantly more A responses than participants in a prevention focus ($M=0.389$, $SD=0.03$) [$t(18)=3.30$, $p<.01$]. For participants in the losses reward structure, participants in a prevention focus ($M=0.522$, $SD=0.03$) made significantly more A responses than participants in a promotion focus ($M=0.330$, $SD=0.01$) [$t(18)=5.97$, $p<.001$].

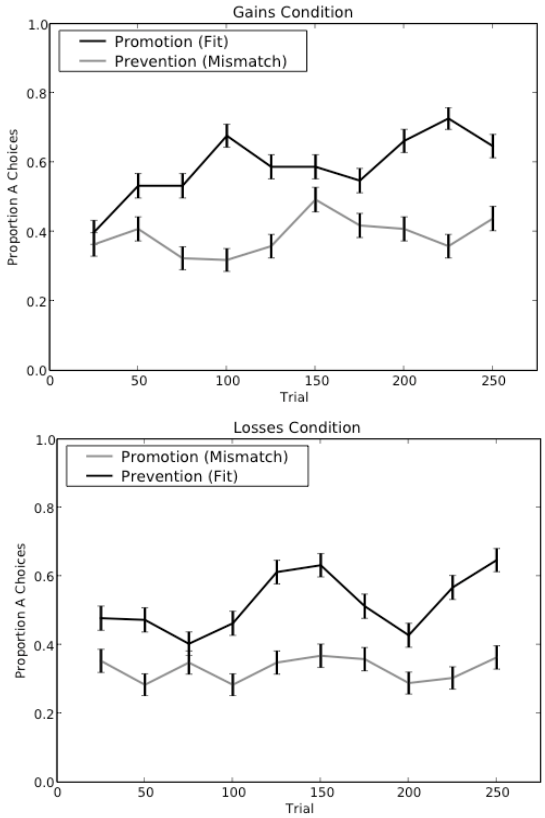

Figure 3: Proportion optimal (A) choices made, by group, over the course of the 250 trials, in bins of 50 trials.

Figure 3 shows the proportion of "A" choices calculated over blocks of 25 trials at a time averaged across participants. We conducted a 2 (regulatory focus) x 2 (reward structure) x 5 (trial block) ANOVA on number of A choices made across the course of the 5 blocks, revealing a significant 2-way interaction between regulatory focus and reward structure ($F(1,38)$ =17.32, $p<.001$), as well as a significant main effect of reward structure ($F(1,38)$ = 4.48, $p<.05$) and a significant main effect of trial block ($F(1,38)$ = 7.86, $p<.01$). All other main effects and interactions failed to reach significance.

As another measure of optimal performance, we calculated each participant's final distance from the bonus criterion, as depicted in Figure 4. A 2 (regulatory focus) x 2 (reward structure) ANOVA on this measure revealed a significant interaction ($F(1,38)=20.05$, $p<.001$) and no significant main effects. Among participants in the gains reward structure, participants in a promotion focus (M= 39.96, SD=8.67) came significantly closer to the bonus criterion than did participants in a prevention focus (M= 66.53, SD= 4.91) [$t(18)=2.66$, $p<.05$]. For participants in the losses reward structure, participants in a prevention focus (M=52.68, SD= 4.40) ended significantly closer to the bonus criterion than participants in a promotion focus (M= 75.06, SD= 0.72) [$t(18)=5.022$, $p<.001$].

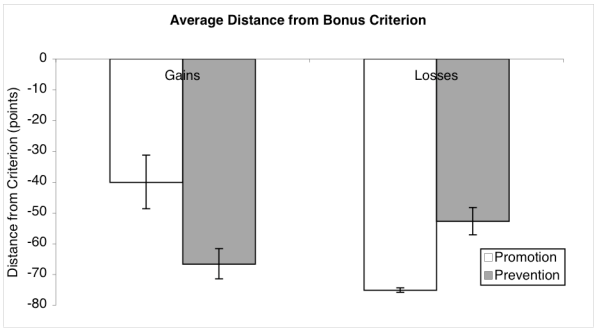

Figure 3: Average distance, in points, from bonus criterion by group.

Model-Based Analysis

Model Definition We implemented a variant of temporal difference (TD)-learning known as average reward learning (Schwartz, 1993; Sutton & Barto, 1998), as theoretical work suggests that average-reward may be a more realistic model of human behavior than discounted-reward models (Daw & Touretzky, 2000; Gureckis & Love, in press). Our descriptive model affords a direct assessment of a given participant's reactivity to local perturbations in payoffs in the rising optimum task.

Unlike standard TD-learning RL models (e.g. Yechaim & Busemeyer, 2005), which rely only on estimated values of individual actions, average reward learning maintains an estimate of the average reward per time step, ρ, across both actions. The value of an action is defined by its estimated value *relative* to the average reward. Thus, actions that lead to better-than-average rewards (i.e., positive transient differences with respect to ρ) are selected more frequently under an exploitative policy. Under average reward learning, the TD error δ is defined as:

$$\delta = r_{t+1} - \rho - Q(a_i)\ , \qquad (1)$$

where r_{t+1} is the actual experienced reward on that trial and ρ is the model's average reward per time step estimate. Each trial, the update made to the estimated transient value, $Q(a_j)$ of each action a_j is informed by the current TD error δ:

$$Q(a_j) = Q(a_j) + \alpha \cdot e_j \cdot \delta, \qquad (2)$$

where α is a learning rate parameter, $0 \le \alpha \le 1$ and e_j is an eligibility trace for that action (described below). If the chosen option a_i had the greater estimated value between the two choices, the average reward estimate ρ is updated according to the current TD error δ:

$$\rho = \rho + (\beta \cdot \delta), \qquad (3)$$

where β is an average-reward update size parameter, $0 \le \beta \le 1$, that determines how heavily the average-reward estimate weights recent rewards. When β is small, ρ relies on a large historical window and updates very slowly, while if $\beta = 1$, ρ depends only on rewards from the most recent trials and is updated quickly. Thus, a participant's readiness to update their expectations of average, trial-to-trial payoffs could be encapsulated by their average-reward update parameter.

Finally, the model utilizes the "softmax" method of action selection whereby the probability of making choice a_i each trial is:

$$P(a_i) = \frac{\exp[\gamma \cdot Q(a_i)]}{\sum_{j=1}^{2} \exp[\gamma \cdot Q(a_j)]} \quad (4)$$

where γ is an exploitation parameter (c.f., Daw et al., 2006; Sutton & Barto, 1998) and $Q(a_i)$ is an estimate of the transient reward associated with choice a_i.

In order to effectively manage temporal credit assignment—that is, the rewarding or penalizing of past choices which occur at variable times prior to the current reward—the model utilizes accumulating eligibility traces in a manner similar to Bogacz et al. (2007), as shown in Equation 2. The eligibility trace e_j for each action is initialized to 0 at the start of the trials and after each action, both eligibility traces are decayed by a constant term and the eligibility trace for the chosen action e_i is incremented:

$$e_j = \lambda e_j \quad (5)$$

$$e_i = e_i + 1 \quad (6)$$

where λ is a decay parameter, $0 \leq \lambda \leq 1$. Eligibility traces improve the rate of learning by allowing prediction errors to propagate backwards across multiple trials (Sutton & Barto, 1998).

Model Fit Predictions and Results Consider a decision-maker, who passes the crossing point from left to right (see Figure 1) as they continually make A choices. If the decision-maker readily changes their average-reward estimate (i.e., large β) to reflect the dip in payoffs encountered, they will seek the high positive transient obtained from choosing B and move back towards the "matching" crossing point, maintaining a suboptimal choice allocation. However, if the decision-maker does not significantly change their estimate (i.e., small β) as they depart from the crossing point, their average-reward estimate will remain anchored roughly at the crossing point, meaning that choice B will not incur as large a transient payoff as it would if the average-reward estimate followed the dip. Consequently, choices A and B will have closer estimated transient values, and thus, will be more equiprobable choices under softmax action selection. Thus, in a sense, slower average-reward updating makes exploration tenable from the perspective of the local decision-maker.

The examination of group differences with respect to average-reward update size parameter (β) values would allow us to evaluate the degree to which participants' expectancies of global payoffs fluctuate with changes in local payoffs. As our behavioral results suggested that regulatory fit affected participants' levels of reactivity to local payoff changes, we hypothesized that participants in a regulatory fit would be slower to update their expectations of average per-trial payoffs, and thus yield lower estimates of the average-reward update size parameter than would participants in a regulatory mismatch. We fit this model to the data using a parameter optimization procedure that maximized the likelihood of the each individual participant's estimated parameter values given their choice behavior over 250 trials (see Yechaim & Busemeyer (2005) for details). To ensure our average-reward model captured participants' response dynamics, we also fit a single-parameter baseline model to each participant's data, which assumed a constant probability of making A choices across all trials. The proportions of subjects in each condition for whom the average-reward model provided a better fit than the baseline model (by the Akaike Information Criterion, see Akaike, 1974) are reported in Table 2.

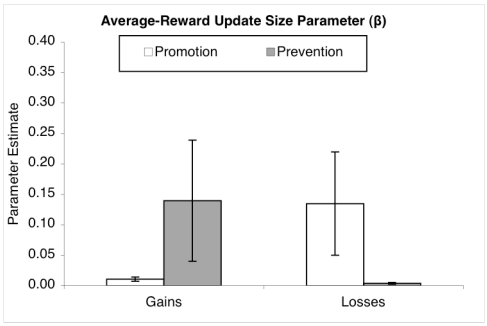

Figure 5: Average estimates of average-reward update size parameter β by condition.

Figure 5 depicts the average update size parameter values for each condition. A 2 (regulatory focus) x 2 (reward structure) ANOVA conducted on estimated update size parameters revealed a significant interaction ($F(1,38)= 6.56$, $p<.05$). On average, participants in regulatory fit had lower estimated values of this parameter. The estimated values for the four model parameters are also summarized in Table 2. The estimated values of γ, α, and λ were not of interest in this analysis, and no significant interactions or main effects were found across the four conditions.

Table 2: Proportion of Subjects for whom Average-Reward Fit Best, and Average Estimated Parameter Values by Experimental Condition. Standard Deviations for these Parameter Values are Shown in Parentheses.

Condition	*Proportion Best Fit*	γ	α	β	λ
Promotion-Gains	0.70	17.839 (6.797)	0.061 (0.094)	0.010 (0.011)	0.559 (0.878)
Promotion-Losses	0.70	18.478 (6.884)	0.034 (0.039)	0.134 (0.268)	0.540 (0.597)
Prevention-Gains	0.80	11.547 (7.872)	0.192 (0.287)	0.139 (0.315)	0.514 (1.02)
Prevention-Losses	0.80	15.267 (9.660)	0.108 (0.194)	0.004 (0.005)	0.567 (0.803)

Discussion

This report examines the effects of regulatory fit on optimal decision-making performance in a dynamic task environment. While previous research has addressed the neural correlates of "risky" choice behavior (Montague & Berns, 2002) and the effects of decaying memory for actions between choices (Bogacz et al., 2007) in decision-making environments where payoffs vary as a function of recent behavior, little work has examined motivational factors that bear on performance in this class of tasks. We have shown

that regulatory fit strongly influences how human choice behavior adapts to changing payoff contingencies in the environment. Specifically, we revealed that compatibility between one's situational regulatory focus and the reward structure of the environment diminishes one's reactivity to local changes in payoffs—which, in the rising optimum task, is necessary for optimal, payoff-maximizing patterns of choice. It should be noted, however, that optimal choice behavior did not depend solely on the reward structure of the environment (e.g., gains and losses), but rather the interaction between situational regulatory focus and task reward structure.

A possible interpretation of differential levels of sensitivity to local payoff changes is that continually modifying one's response policy on the basis of local payoff information impedes *systematic exploration* of the decision space. That is, reactivity to local changes in payoffs precludes full, systematic exploration of the decision space. The notion of systematic exploration is closely related to "temporal abstraction" in reinforcement-learning as described by Botvinick et al. (in press) by which agents can reduce the effective size of the decision space through structured, multiple-action patterns of exploration. While previous accounts of motivational influences of choice in bandit tasks find that regulatory fit engenders more stochastic decision-making on the independent, trial-to-trial level (Worthy et al., 2007), participants' choice behavior in the present work suggests that regulatory fit also facilitates a more systematic form of exploration which persists over multiple choices.

We have shown in this report that motivational factors in the environment can influence individuals' level of reactivity to local payoff changes in a dynamic decision-making task, which can in turn impact their willingness to explore globally optimal choice strategies. These results add to the body of findings from the decision-making and classification literatures (Maddox, Baldwin, & Markman, 2006; Worthy et al., 2007), which suggest motivation holds strong effects for human cognition and behavior.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control 19(6),* 716-723.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50(1-3),* 7-15.

Bogacz, R., McClure, S. M., Li, J., Cohen, J. D., & Montague, P. R. (2007). Short-term memory traces for action bias in human reinforcement learning. *Brain Research, 1153,* 111-21.

Botvinick, M. M., Niv, Y., & Barto, A. C. (in press). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition.*

Carver, C.S. & Scheier, M.F. (1998). *On the Self-Regulation of Behavior*. New York: Cambridge University Press.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. Nature, 441, 876-879.

Daw, N. D., & Touretzky, D. S. (2000). Behavioral considerations suggest an average reward TD model of the dopamine system. *Neurocomputing, 32-33,* 679-684.

Eagelman, D. M., Person, C., & Montague, P. R. (1998). A Computational Role for Dopamine Delivery in Human Decision-Making. *Journal of Cognitive Neuroscience, 10(5),* 623-630.

Gureckis, T.M., & Love, B.C. (in press) Learning in noise: dynamic decision-making in a variable environment. *Journal of Mathematical Psychology.*

Herrnstein, R. J. (1990). Rational choice theory: Necessary but not sufficient. *American Psychologist, 45(3),* 356-367.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52(12),* 1280-300.

Maddox, W. T., Baldwin, G. C., & Markman, A. B. (2006). A test of the regulatory fit hypothesis in perceptual classification learning. *Memory & Cognition, 34(7),* 1377-97.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36(2),* 265-84.

Otto, A.R., Gureckis, T.M., Markman, A.B., & Love, B.C. Navigating through Abstract Decision Spaces: Evaluating the Role of State Generalization in a Dynamic Decision-Making Task. Submitted.

Schwartz, A. (1993). A reinforcement learning method for maximizing undiscounted rewards. In *Proceedings of the Tenth International Conference on Machine Learning,* 298-305. Amherst: Morgan Kaufmann.

Shah, J. & Higgins, E.T. (1997). Expectancy * value effects: Regulatory focus as determinant of magnitude and direction. *Journal of Personality and Social Psychology, 73 (3),* 447–58.

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

Worthy, D. A., Maddox, W. T., & Markman, A. B. (2007). Regulatory fit effects in a choice task. *Psychonomic Bulletin & Review, 14(6),* 1125-32.

Yechiam, E., & Busemeyer, J. R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin & Review, 12(3),* 387-402.

Yerkes, R. M., & Dodson, J. D. (1908) The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.

Acknowledgments

This research was supported in part by AFOSR Grant FA9550-07-1-0178 and NSF CAREER Grant 349101 to Bradley C. Love, and NIMH Grant MH077708 and AFOSR Grant FA9550-06-1-0204 to W. Todd Maddox and Arthur B. Markman.