Journal of Experimental Psychology:
Learning, Memory, and Cognition
2010, Vol. 36, No. 3, 797–804

0278-7393/10/$12.00 DOI: 10.1037/a0018999

Regulatory Fit and Systematic Exploration in a Dynamic Decision-Making Environment

A. Ross Otto and Arthur B. Markman
University of Texas at Austin

Todd M. Gureckis
New York University

Bradley C. Love
University of Texas at Austin

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 that participants in a regulatory fit exhibit increased levels of exploratory choice and flexible use of multiple strategies over the course of an experiment. The present study placed promotion and prevention-focused participants in a dynamic environment for which optimal performance is facilitated by systematic exploration of the decision space. These participants either gained or lost points with each choice. Our experiment revealed that participants in a regulatory fit were more likely to engage in systematic exploration of the task environment than were participants in a regulatory mismatch and performed more optimally as a result. Implications for contemporary models of human reinforcement learning are discussed.

Keywords: decision making, reinforcement learning, exploration, motivation, regulatory fit

A central assumption of contemporary work in learning and decision making is that rational agents attempt to maximize reward. However, the motivation literature highlights how the interpretation of feedback or rewards is strongly influenced by active goals held by the learner. In particular, a distinction is made between approach goals—desirable end states that one wants to work toward—and avoidance goals—undesirable end states that one wishes to prevent from occurring (Carver & Scheier, 1998). As an example, consider the contrast between working toward earning an A grade versus working to avoid failure of a class. Intuitively, it is unlikely that equivalent psychological states would result from these two situations. Higgins (1997) suggested that two distinct psychological states, evoked by approach and avoidance goals, tune the sensitivity of the motivational system to gains and losses in the environment.

Under this view, a promotion focus activates an approach mode of processing that focuses the motivational system on the presence or absence of gains, while a prevention focus activates an avoidance mode of processing, focusing the motivational system on the presence or absence of losses. There is a tendency for people to have a chronic regulatory focus, but situations often induce a regulatory focus that can overpower this long-term tendency (Shah, Higgins, & Friedman, 1998).

Recent research suggests that the influence of regulatory focus on cognition depends on the interaction between one's regulatory focus and the local reward structure of the environment—that is, gains or losses resulting from one's actions (Higgins, 2000; Maddox, Baldwin, & Markman, 2006). When one's regulatory focus matches the environment's reward structure, it is called a *regulatory fit*. Conversely, when regulatory focus and reward structure do not match, it is called a *regulatory mismatch* (see Figure 1).

This line of research suggests that regulatory fit appears to promote more flexible cognitive strategies across a number of dissimilar tasks. For example, Worthy, Maddox, and Markman (2007) examined decision making in a bandit task for which optimal choice behavior required balancing exploration of unknown options against the desire to exploit options that are known to be rewarding. Participants given a promotion focus by attempting to earn a prize made more exploratory choices (i.e., those that did not maximize expected utility) when they tried to maximize the points gained from the decks of cards (i.e., regulatory fit) than when they tried to minimize losses (i.e., regulatory mismatch). In contrast, participants given a prevention focus by trying to avoid losing a prize were more exploratory when they tried to minimize points lost (regulatory fit) rather than when they tried maximizing points gained (regulatory mismatch). The implication is that regulatory fit leads to more flexible or exploratory pattern of behavior than does regulatory mismatch (a pattern also found in other tasks such as category learning; Maddox et al., 2006). However, the exact mechanism by which regulatory fit influences exploratory behavior is not clear. In this report, we consider the notion of

A. Ross Otto, Arthur B. Markman, and Bradley C. Love, Department of Psychology, University of Texas at Austin; Todd M. Gureckis, Department of Psychology, New York University.

This research was supported in part by Air Force Office of Scientific Research Grant FA9550-07-1-0178 to Bradley C. Love, and National Institute of Mental Health Grant MH077708 and Air Force Office of Scientific Research Grant FA9550-06-1-0204 to W. Todd Maddox and Arthur B. Markman.

Correspondence concerning this article should be addressed to A. Ross Otto, Department of Psychology, University of Texas, Austin, TX 78712. E-mail: rotto@mail.utexas.edu

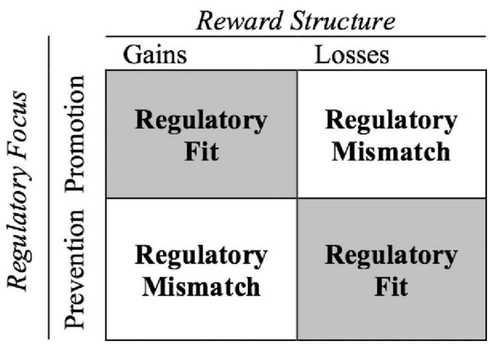

Figure 1. Overview of states evoking regulatory fit and mismatch. When one's situational regulatory focus matches the reward structure of the environment, a regulatory fit results. In contrast, when one's situational regulatory focus does not mach the reward structure of the environment, a regulatory mismatch results.

exploration more deeply to shed light on the processes decision makers may use to make optimal choices.

In typical tasks used to study reward learning—such as the bandit problem used in the Worthy et al. (2007) experiment—the payoffs of each option are independent of the choices just made by the participant (e.g., Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). Theoretical accounts of how people balance exploitation of known information with exploration in order to collect new information are a recent focus of study in the area of neural reinforcement learning (RL). Computationally, participants' tendencies toward exploration or exploitation are often characterized by the softmax rule, which parameterizes the tendency to explore relative to the tendency to exploit (Sutton & Barto, 1998). Worthy et al. found that participants in a fit exhibited increased exploration of choices—with the unavoidable consequence of occasionally choosing against their best interest. Model-based analyses revealed that participants in a regulatory fit made more stochastic decisions than did participants in a regulatory mismatch. Quantitatively, participants in a regulatory fit were found to have smaller values of the softmax exploitation parameter than did participants in a regulatory mismatch, providing a formal framework for understanding the impact that regulatory fit has on exploratory choice behavior.

However, the effectiveness of particular exploratory behaviors hinges on the nature of the decision-making environment. Consider foraging in a physical space, where one's location can be thought of as the current state. Assume that the most rewarding state is far away from the decision maker's current state. To reach this state, the decision maker must have a desire to explore distant novel states and discover the rewards in these states. To effectively reach these distant states to discover their rewards, the decision maker must repeatedly produce the same action (e.g., moving west), as undirected random movement will likely not get one very far in any direction. Consider, as an anecdote, if Christopher Columbus's exploration was defined by merely sailing in a random direction each day: His ships would not likely have traveled a few miles beyond the harbor! Thus, efficient exploration entails dependencies between one's current choice and past choices.

By contrast, in the softmax framework, exploration is characterized as a stochastic choice made independently of previous choices. In light of this distinction, it is not clear whether a regulatory fit engenders more systematic exploration of decision spaces—whereby exploration is a multiple-trial phenomenon (akin to "temporal abstraction"; Botvinick, Niv, & Barto, 2009)—or whether a regulatory fit simply results in more stochastic, trial-independent choices. The bandit task used by Worthy et al. (2007) does not disambiguate between these two methods of exploration, because both methods support effective learning of choice payoff contingencies. In this article, we examine choice behavior in a task that helps us disentangle these possibilities. This distinction is theoretically significant because it helps determine whether a regulatory fit simply increases the "noise" in individuals' response strategies (which helps in some environments) or whether it brings about a more systematic change in the way people explore their environment. Previous investigators have found that a regulatory fit also facilitates generation of anagram solutions (Shah et al., 1998) and generation of correct associates for difficult remote associates task (Mednick & Mednick, 1967) items (Markman, Maddox, & Baldwin, 2007). These tasks do not elucidate the nature of exploratory behavior brought about by regulatory fit.

The task we use is a two-option, repeated-choice, decision-making task termed the *rising optimum*, previously used to investigate simple RL accounts of behavior (Montague & Berns, 2002) and the problem of temporal credit assignment in human sequential decision making (Bogacz et al., 2007). The payoff structure for this task is illustrated in Figure 2. Unlike standard bandit tasks, payoffs on each trial are dependent on the proportion of selections made to each option over a 20-trial moving window. Thus, this proportion of responses defines the current state of the task environment (Gureckis & Love, 2009). If the next response changes the relative proportion of the previous 20 responses, then the state changes. In the rising optimum, there are 21 possible states.

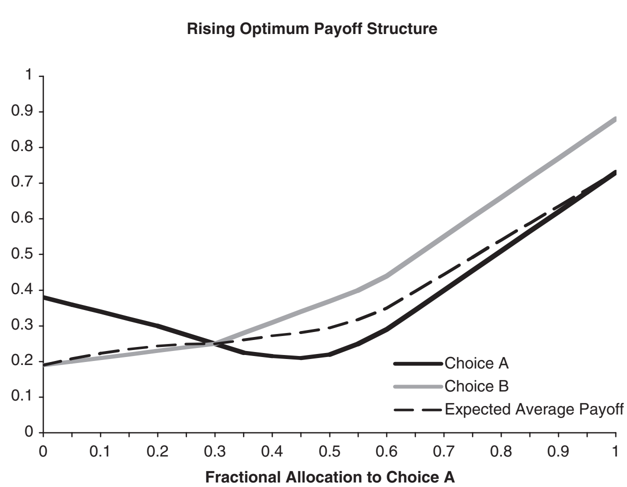

Figure 2. Payoff functions for two choices as a function of choice allocation. The payoff received from a choice depends on the proportion of A choices made over the last 20 trials, represented by the horizontal axis. The solid black and gray lines depict the payoff curves for Choice A and Choice B, respectively, and the dashed line depicts the expected average payoff for each choice allocation. In this task, optimal choice behavior requires consistent A choices every trial.

For example, if the participant makes only B choices for 20 consecutive trials—effectively making his or her fractional allocation to the A choice 0—the payoffs from an additional selection of Choice A or B would be .38 and .19, respectively. If the participant makes one A choice at this point, his or her response allocation would change to .05, as only one out of 20 of the last trials were A choices. Consequently, the payoffs for the next choice of A or B would be .36 and .20, respectively. Thus, the payoffs associated with the each choice fluctuate with the past choice behavior of the participant. It is important to note that a choice remains in the participant's choice history for 20 trials—with the consequence that a single B choice will prevent a participant from reaching an allocation of one for 20 trials. In this task, optimal long-term choice behavior requires consistent A choices as the global optimum is located where the participant's fractional choice allocation to choice A is one.

Prior experimental work 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 .30 (Bogacz et al., 2007). To illustrate, consider participants who make repeated A choices until they find themselves at the crossing point of the two curves (see Figure 2). As they continue to make A choices and move rightward, they will encounter a dip in payoffs, potentially discouraging further movement in that direction. At this point, greater payoffs resulting from B can lure the decision makers leftwards, until they pass the crossing point and A becomes more attractive again. This globally suboptimal response strategy is predicted by simple RL models of choice such as the temporal-difference learning algorithm (Montague & Berns, 2002). In this model, decision making is driven by individual outcomes at the local, trial-to-trial level, which results in suboptimal behavior.

In the rising optimum task, finding the globally optimal strategy requires exploration of the possible states in the decision space. However, unlike the bandit tasks used in previous research, this task structure requires more systematic behavior to explore the domain. For example, in order to move to the right of the crossing point, participants must make repeated selections to one of the options (e.g., five B choices followed by 15 A choices). Thus, the task is more akin to the foraging example above, in that a decision maker is more likely to uncover potentially rewarding states if he or she makes the same actions repeatedly. Systematic responses like this are hard to model with the softmax rule, because this rule assumes that the decision made on each trial is probabilistic and is independent of the choices made on previous trials. While decision makers utilizing an entirely exploitative (i.e., deterministic) decision rule will find themselves cycling around the crossing point of the payoff curves (Montague & Berns, 2002), a stochastic rule will lead decision makers to explore the state space in a random walk fashion. In the limit, as the number of trials goes to infinity, the stochastic softmax decision rule will eventually visit every state in the task. However, this undirected exploration is unlikely to visit every state in a small number of trials. As an illustration, the solid line in Figure 3 presents a single simulation of a softmax model with a stochastic decision rule. This model fails to explore far beyond the middle of the state space. Exploring in a more systematic fashion by making long streaks of selections to single choices allows learners to traverse the state space more effectively and uncover the optimal choice allocation. A simulation of a softmax model augmented with a parameter governing the model's tendency to repeat individual choices (described in detail below) is depicted by the dotted line in Figure 3. Streaky choice behavior leads to a broader exploration of the state space.

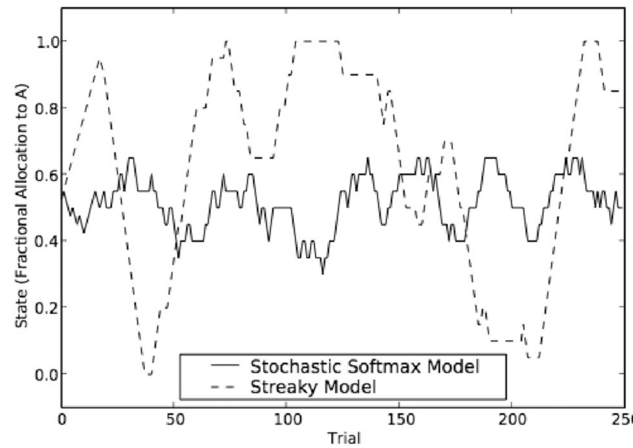

Figure 3. Example of two individual model simulations of the rising optimum task: a softmax model with a stochastic decision rule (solid line) and a streaky decision rule mode (dashed line).

In this study, we teased apart these exploration processes in order to better understand how regulatory fit bears on the decision maker's tendency to engage in systematic exploration of the environment. To this end, we manipulated participants' global motivational state (i.e., regulatory focus) and the environment reward structure to determine the influence of regulatory fit on choice behavior in the rising optimum task. We reasoned that if a regulatory fit leads to more stochastic responding than a regulatory mismatch, then we would expect that neither participants in a regulatory fit nor in a regulatory mismatch would find the globally optimal strategy in this task (i.e., 100% allocation to A). Alternatively, if regulatory fit facilitates systematic exploration of the task decision space, we expected that participants in a regulatory fit should be more likely than participants in a regulatory mismatch to find the globally optimal strategy. According to both of these accounts, we reasoned that regulatory mismatch participants should generally make decisions informed by local, trial-to-trial estimates of choice payoffs in a manner consistent with simple softmax accounts of learning in the rising optimum task (Montague & Berns, 2002), resulting in overall choice allocations around the crossing point.

To maximize payoffs in the rising optimum task, the learner must systematically explore in order to uncover the possible states in the decision space, a strategy that is not apparent at the outset of the task. We have reason to believe that regulatory fit should promote systematic exploration in the rising optimum task. Decision makers in a regulatory fit have been shown to continually sample options in bandit tasks in order to gather more information about potential changes in payoffs, whereas decision makers in a regulatory mismatch exploit their early knowledge about payoffs and fixate on an early solution (Worthy et al., 2007). More abstractly, this work suggests that a regulatory fit engenders the use of strategies that reduce uncertainty about rewards in the environment. In the rising optimum task, we expected that regulatory fit drives choice behavior toward exploring distant states with unknown rewards.

We hypothesized that participants in a regulatory fit would seek out all possible states of the decision space in order to uncover the optimal choice allocation (100% A choices). To this end, they should engage in systematic exploration, exhibiting streaks of single choices in order to move about the decision space. In contrast, we hypothesized that participants in a regulatory mismatch would not make choices directed by the reduction of uncertainty in distant states, and thus would exhibit suboptimal local matching behavior (i.e., choice allocations near the crossing point). Note that our expected behavioral effects do not rely on regulatory focus or reward structure singularly but rather on their crossover interaction as in Maddox et al. (2006) and Worthy et al. (2007). Besides contributing to a deeper understanding of the influence of regulatory fit on exploratory choice, understanding motivational factors that influence effective learning in environments with uncertain structures at the outset is of practical concern.

Experiment

We placed participants in a variant of the rising optimum task, whose payoff schedule (under the gains reward structure) is depicted in Figure 2. The gains and losses framing manipulations were designed to result in informationally equivalent situations (Shah et al., 1998). 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 criterion was positioned such that participants would need to earn at least 75% of the total possible points at the end of the experiment. Consequently, participants who discovered and persisted with the globally optimal response strategy would meet the bonus criterion, while those who remained near the crossing point would not.

Participants in a promotion focus were told that they would receive an entry into a drawing for a one in 10 chance at winning $50 if their performance met the bonus criterion. Participants in a prevention focus were given an entry into the same drawing and were told that they had to meet the bonus criterion to avoid losing it. This framing manipulation was designed so that participants in the promotion and prevention focus conditions were effectively in the same economic situation (Shah et al., 1998).

Method

Participants. Forty undergraduates from the University of Texas at Austin 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. Participants sat at a computer to perform this study. At the start of the experiment, participants were informed that they would either earn (promotion condition) or keep (prevention condition) an 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 .50 across all participants.[1] For 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 4. For each trial, participants clicked one of the buttons to indicate their choice, and the white payoff bar grew (or fell, in the losses condition) vertically to indicate the amount of points gained (or lost, in the losses condition). There was no time limit for making choices.

The payoff on 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 2. 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 .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. If they met the bonus criterion, participants in the promotion focus condition were given a ticket and were 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 primary measure of performance, Figure 5 shows the mean proportion of optimal A choices calculated over blocks of 25 trials. The results were consistent with the hypothesis that participants in a regulatory fit were better able to find the globally optimal choice strategy. We conducted a 2 (regulatory focus) × 2 (reward structure) × 10 (trial block) analysis of variance (ANOVA) on the number of A choices made across the 10 blocks, revealing a significant two-way interaction

[1] To establish that differences in choice performance were not due to starting choice allocations (which were randomly determined), we calculated the correlation between initial allocation to A and overall choice allocations across participants and found no significant relationship, $r(38) = -.07$, $p = .67$. Additionally, a 2 (regulatory focus) × 2 (reward structure) ANOVA conducted on initial choice allocations did not reveal a significant interaction, $F(1, 38) = 0.49$, $p = .49$, or main effects of reward structure, $F(1, 38) = 0.34$, $p = .56$, and regulatory focus, $F(1, 38) = 1.04$, $p = .31$. In summary, we found little reason to believe that starting allocations influenced overall choice behavior or differed systematically between groups.

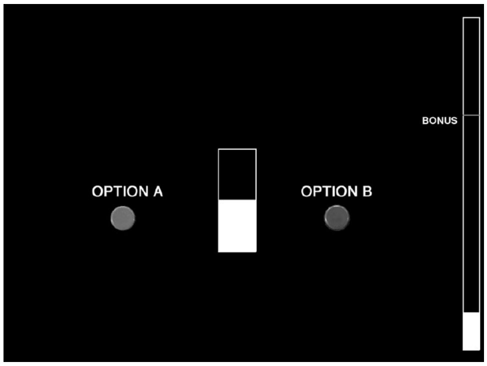

Figure 4. The task interface shows the two choice buttons, the rising/falling payoffs bar in the center, and the cumulative point meter (in relation to the bonus criterion) on the right side of the display.

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, $F(9, 30) = 7.86$, $p < .01$. 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 (i.e., points short of the bonus criterion at the end of 250 trials), depicted in Figure 6. A 2 (regulatory focus) × 2 (reward structure) ANOVA on this measure revealed a significant interaction, $F(1, 38) = 20.05$, $p < .001$, with 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 did participants in a promotion focus ($M = 75.06$, $SD = 0.72$), $t(18) = 5.02$, $p < .001$.

We also analyzed the overall proportion of trials for which participants made optimal A choices, collapsed over the course of the experiment. A 2 (regulatory focus) × 2 (reward structure) ANOVA 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 did 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 did participants in a promotion focus ($M = 0.330$, $SD = 0.01$), $t(18) = 5.97$, $p < .001$.

Model-based analysis. While the performance results suggest that regulatory fit and mismatch affect the degree to which participants optimally allocate choices, they do not directly address our prediction that participants in a regulatory fit would exhibit more systematic exploration of the decision space than would participants in a regulatory mismatch. The aim of our model-based analysis was to illuminate patterns of systematic exploration in participants' choice behavior using choice streakiness as a proxy measure.

The two-parameter standard softmax model predicts the probability of making choice a_i on trial t, informed by the participant's choice and outcome experience up to trial t:

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

where γ is an exploitation parameter (Sutton & Barto, 1998) and $Q(a_i, t)$ is an estimate of the payoff associated with choice a_i. The model utilizes an incremental updating rule to inform estimated choice payoffs at the next trial $t + 1$:

$$Q(a_j, t + 1) = Q(a_j, t) + \alpha[r(t + 1) - Q(a_j, t)], \tag{2}$$

where α is a learning rate parameter, $0 \leq \alpha \leq 1$, and $r(t + 1)$ is the payoff from the chosen option a_j. This model is identical to that used by Worthy et al. (2007).

The extended model adds a third streak parameter, p_{repeat}, that specifies the probability with which the model makes the same choice on trial $t + 1$ as it did on trial t. The probability of choosing

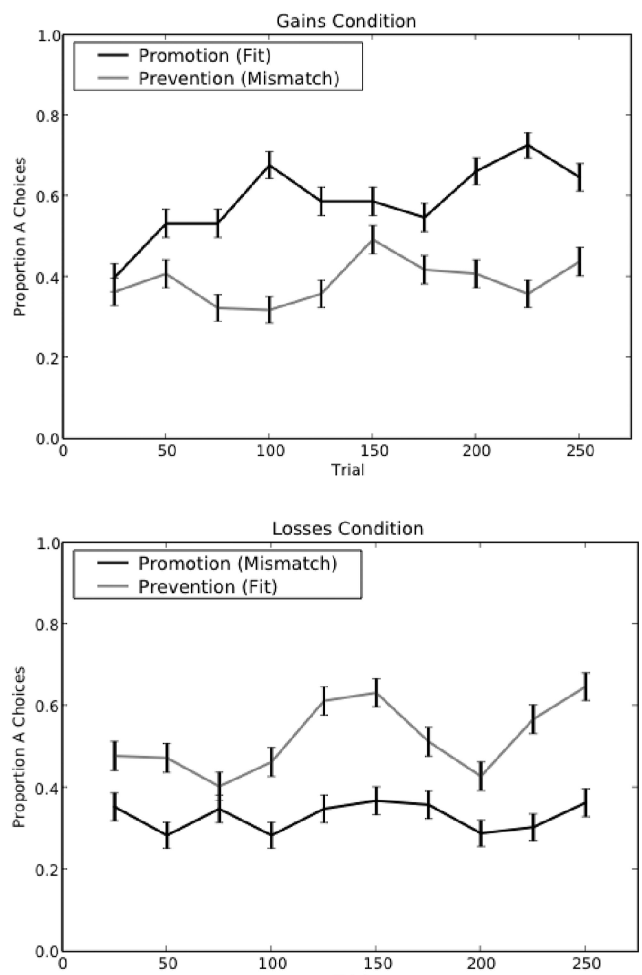

Figure 5. Proportion optimal (A) choices made, by group, over the course of the 250 trials, in blocks of 50 trials. Error bars indicate standard error of the mean.

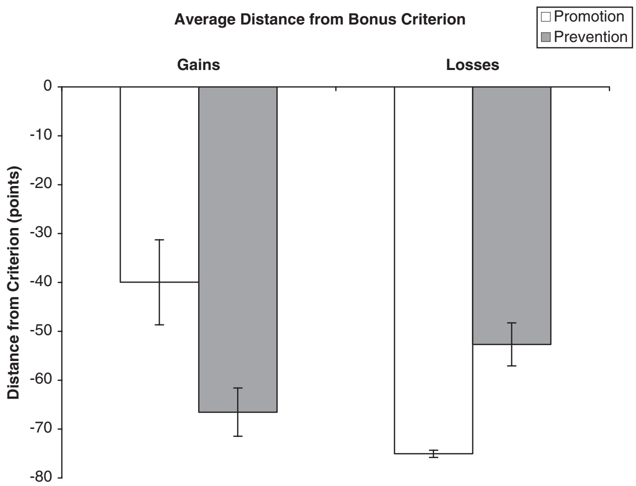

Figure 6. Average distance, in points, from bonus criterion by group. Distances are equated for gains and losses payoff structures. Error bars indicate standard error of the mean.

action a_i is computed from the softmax probability from Equation 1 and p_{repeat}:

$$P(a_i, t|choice_{t-1} = a_i)$$

$$= p_{repeat} + (1 - p_{repeat})\left(\frac{\exp[\gamma \cdot Q(a_i, t)]}{\sum_{j=1}^{2} \exp[\gamma \cdot Q(a_j, t)]}\right). \quad (3)$$

Intuitively, larger values of p_{repeat} indicate larger streaks of selections of the same option. For each model, we sought parameter estimates that maximized the likelihood of each participant's observed choices:

$$L_{model} = \prod_t P_{c,t} \quad (4)$$

where c_t reflects the choice made on trial t, informed by participant's choice and payoff experience up to trial t. We used G^2 to assess the fit of the streaks model relative to the standard nested softmax model, where

$$G^2 = 2[\ln(L_{streak}) - \ln(L_{\mathrm{softmax}})]. \quad (5)$$

Figure 7 depicts the proportion of participants, by condition, best fit by the streak model (relative to the chi-square distribution at $\alpha = .05$, $df = 1$). A three-way chi-square test revealed that best-fitting model was not independent from motivational condition, $\chi^2(2, N = 104) = 18$, $p < .05$, suggesting that regulatory fit influenced the degree to which the streak parameter improved the fit of the standard softmax model.

As expected, we found that people in a regulatory fit were streakier in their choices than were those in a regulatory mismatch. It is possible that people in a fit were more likely to find the optimal state and made repeated choices of A in order to maintain that state. In this case, the streaky model could simply be redescribing stable optimal choice behavior. To disentangle these possibilities, we compared goodness of fit of the standard and streaky models for only exploratory choices—that is, trials in which participants selected options with smaller estimated payoffs (cf. Daw et al., 2006). A better fit by the streaky model would suggest response streaks were being made to traverse the decision space in the face of inferior immediate payoffs. Ninety percent of participants who were best fit overall by the streaky model were also best fit by the streaky model for exploratory trials (G^2 relative to the chi-square distribution at $\alpha = .05$, $df = 1$), suggesting that the streaky model fits are describing exploratory choice behavior and not merely redescribing stable optimal choice behavior.

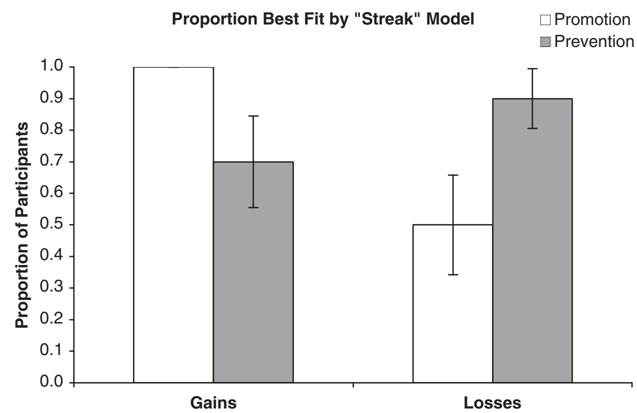

Figure 7. Proportion participants best fit by streaks model, by condition. Error bars indicate standard error of proportion.

Further, we analyzed the estimated values of the streak parameter across all participants regardless of the model that best characterized their choice behavior, in order to evaluate the degree to which motivational condition affected participants' tendency to make long streaks of single choices. The average estimated values of this parameter are depicted in Figure 8. A 2 (regulatory focus) × 2 (reward structure) ANOVA revealed a significant interaction between regulatory focus and reward structure, $F(1, 38) = 9.38$, $p < .01$. In the gains reward structure, participants in a promotion

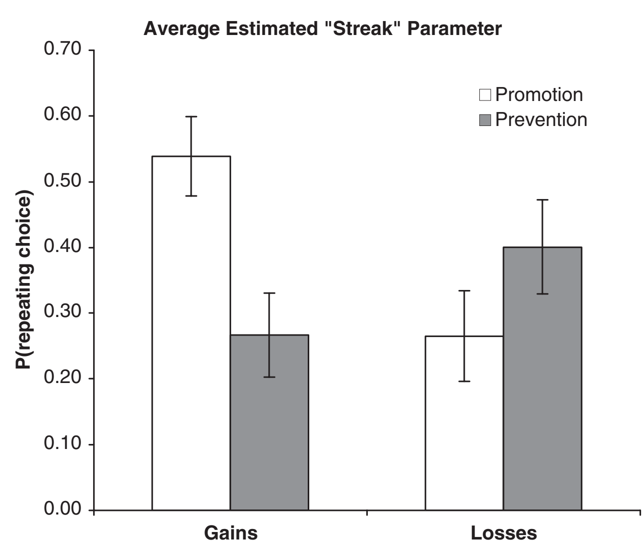

Figure 8. Average estimates of streakiness parameter p_{repeat}, by motivational condition. Error bars indicate standard error of the mean.

focus ($M = 0.54$, $SD = 0.19$) exhibited higher values of the streak parameter than did participants in a prevention focus ($M = 0.27$, $SD = 0.20$), $t(18) = 3.41$, $p < .01$. In the losses reward structure, participants in a prevention focus ($M = 0.40$, $SD = 0.23$) exhibited higher values of the streak parameter than did participants in a promotion focus ($M = 0.26$, $SD = 0.22$), although the effect was not significant, $t(18) = 2.13$, $p = .09$. There were no significant main effects of regulatory focus or reward structure. The results of these two model-based analyses suggest that participants in a regulatory fit reveal a signature of streaky behavior, which facilitates systematic exploration in the rising optimum.

Discussion

We examined the effects of regulatory fit on choice in a dynamic task environment in which discovering the reward-maximizing strategy is facilitated by systematic exploration. While previous studies of the effect of motivation on choice behavior have found that regulatory fit engenders more flexible or exploratory decision making, these results have typically been modeled as trial-independent noise (Worthy et al., 2007). The present work suggests that regulatory fit also facilitates a more systematic form of exploration that persists across multiple trials. In particular, we found that compatibility between regulatory focus and the reward structure of the environment engenders systematic exploration of the decision space, improving the ability to find a reward-maximizing strategy. It should be noted that task performance differences did not depend solely on reward structure (i.e., gains and losses) but rather on the interaction between situational regulatory focus and task reward structure.

Our RL model provides a characterization of systematicity manifested in participants' choice behavior. Note that the streaky tendency described by the model is not a general exploration solution but rather facilitates efficient movement in decision spaces where sequences of repeated responses move the decision maker between states. In our model-based analysis, fits of the streaky model reveal that participants in a regulatory fit are not merely making more stochastic choices as in Worthy et al. (2007) but are exhibiting increased conditional dependency between choices. In the rising optimum, the streak parameter can force the model to make repeated choices of a locally inferior option (e.g., at the dip in the A payoff curve), facilitating efficient traversal of the decision space and discovery of the global optimality of the A option. Specifically, increased streakiness resulted from heightened exploration of the decision space, ultimately leading to better performance.

While the present study augments Worthy et al.'s (2007) empirical account of regulatory fit's influence on exploratory choice, the mechanism by which exploratory behavior arises in these tasks remains an open question. One speculative notion posits that promotion and prevention foci are expectations about the projected state of the world (Markman et al., 2007). Under this view, when the reward structure of the environment fails to match one's expectations, a plausible initial reaction is to engage fast-acting strategies until the environment can be better understood. Conversely, when the reward structure matches expectations, people should bring their full executive resources to bear in that environment. A question arises about the source of the exploratory behavior shown by participants in a regulatory fit in this experiment. It is not yet clear whether regulatory fit, in contrast to regulatory mismatch, promotes a belief that there are many possible states (as in the foraging example above) or whether it merely encourages exploration of distant states. Further, at least two candidate mechanisms could be guiding exploration of distant states. One possibility is that decision makers are initially optimistic about the rewards in to-be-experienced states, driving choice toward these states (so that they have "optimistic initial values": Sutton & Barto, 1998). A second possibility is that choices are directed toward states with higher calculated uncertainty through the use of bonuses for exploration (Dayan & Sejnowski, 1996). These issues should be addressed in future work.

The tendency exhibited by participants in a regulatory fit toward systematic exploration is closely related to "temporal abstraction" in RL by which agents can reduce the effective size of the decision space through structured, multiple-action patterns of exploration (Botvinick et al., 2009). A number of psychological models encapsulate this approach, proposing that individuals make decisions based on the experienced rewards of higher level strategies rather than on individual choices (Erev & Barron, 2005; Gray, Sims, Fu, & Schoelles, 2006; Rieskamp & Otto, 2006). The patterns of choice revealed here suggest the possibility that people may be engaging in more structured exploration in simple bandit tasks than is described by the softmax decision rule. Further, our results underscore the importance of examining the influence of the decision maker's choice history—in addition to reward history—on sequential choice behavior (Lau & Glimcher, 2005). This report adds to the body of findings from the decision-making and classification literatures (Maddox et al., 2006; Worthy et al., 2007) suggesting that motivational factors exert a crucial influence on human cognition and behavior.

References

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,* 111–121.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113,* 262–280.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior.* New York, NY: Cambridge University Press.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006, June 15). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876–879.

Dayan, P., & Sejnowski, T. J. (1996). Exploration bonuses and dual control. *Machine Learning, 25,* 5–22.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931.

Gray, W. D., Sims, C. R., Fu, W., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113,* 461–482.

Gureckis, T. M., & Love, B. C. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition, 113,* 293–313.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52,* 1280–1300.

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist, 55,* 1217–1230.

Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of Experimental Analysis of Behavior, 84,* 555–579.

Maddox, W. T., Baldwin, G. C., & Markman, A. B. (2006). A test of the regulatory fit hypothesis in perceptual classification learning. *Memory & Cognition, 34,* 1377–1397.

Markman, A. B., Maddox, W., & Baldwin, G. C. (2007). Using regulatory focus to explore implicit and explicit processing in concept learning. *Journal of Consciousness Studies, 14,* 132–155.

Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual: Remote Associates Test.* Boston, MA: Houghton Mifflin.

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

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135,* 207–236.

Shah, J., Higgins, E. T., & Friedman, R. S. (1998). Performance incentives and means: How regulatory focus influences goal attainment. *Journal of Personality and Social Psychology, 74,* 285–293.

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,* 1125–1132.

Received May 1, 2009
Revision received December 14, 2009
Accepted December 21, 2009 ■

Correction to Klauer et al. (2010)

In the article "Conditional Reasoning in Context: A Dual-Source Model of Probabilistic Inference," by Karl Christoph Klauer, Sieghard Beller, and Mandy Hütter (*Journal of Experimental Psychology: Learning Memory, and Cognition*, 2010, Vol. 36, No. 2, pp. 298–323), the dual-source model is overparameterized. Only the products $\lambda\tau$ of the λ and τ parameters are uniquely identified by the data. This has no consequences for the ξ parameters, for ratios of τ parameters estimated with the same λ, for ratios of λ parameters associated with the same τ parameters, nor for the fit values. The model fit is, however, achieved more parsimoniously than stated in Klauer et al. because one parameter (Experiments 1, 2, and 4) or two parameters (Experiment 3) are redundant.

To fix the scale for τ and λ parameters, one of them has to be set to one. We recommend to set the largest of τ(MP), τ(MT), τ(AC), and τ(DA) equal to one. This yields unique parameter estimates for τ and λ but has consequences for their interpretation: Differences in overall level of the profile of τ parameters over the four inferences (due to, e.g., differences in cognitive load), if any, would be removed from the τ estimates and would show up in the λ parameters. The above constraint is the one implicitly imposed almost perfectly by the estimation method used in Klauer et al. (2010). In consequence, when the constraint is explicitly enforced, the numerical values of the parameter estimates reported in Klauer et al. change only minimally, and the outcome of all of the significance tests reported remains the same.

DOI: 10.1037/a0019445