Proceedings of the 29th Annual Conference of the Cognitive Science Society
Nashville, Tennessee, USA ❄ August 1-4, 2007

A Computational Model of the Motivation-learning Interface

Manish Saggar (mishu@cs.utexas.edu)
Department of Computer Science, University of Texas at Austin,
Austin, TX 78705 USA

Arthur B Markman (markman@psy.utexas.edu)
W Todd Maddox (maddox@psy.utexas.edu)
Department of Psychology, University of Texas at Austin,
Austin, TX 78705 USA

Risto Miikkulainen (risto@cs.utexas.edu)
Department of Computer Science, University of Texas at Austin,
Austin, TX 78705 USA

Abstract

This paper presents a computational model of how motivation influences learning, elaborating on the empirical study of Markman, Baldwin and Maddox (2005). In a decision criterion learning task with unequal payoffs, the subjects were more likely to maximize the reward when their motivation was in line with the reward structure (i.e. when they were in a regulatory fit), whereas they were more likely to maximize accuracy when their motivation did not match the reward structure (i.e. when they were in a regulatory mismatch). The model accurately replicates this pattern of results, and also accounts for the individual subject's behavior. In addition, the model makes the novel prediction that regulatory-fit subjects who are near the reward threshold will shift their strategy toward maximizing accuracy, whereas regulatory-mismatch subjects who are far from the reward threshold will shift their strategy toward maximizing reward. When the original data was reanalyzed, this model-driven prediction was confirmed. These results constitute a first computational step towards understanding how motivation influences learning and cognition.

Keywords: Computational modeling, Motivation, Learning, Regulatory focus.

Introduction

Many of the tasks in our day-to-day life require us to choose one option over another. Knowingly or unknowingly we make decisions at every step and choose our behaviors from a large repertoire of possibilities. Cognition plays an important role in selection of these behaviors, but our motivational state to approach positive outcomes or avoid negative outcomes also affects which behaviors we select (Maddox, Markman and Baldwin, 2005). Cognitive research has focused on information processing and its effects on learning and behavior with little attention paid to motivation. Recent research in social cognition and cognitive science has begun to bridge this gap (see Maddox, Markman and Baldwin, 2005, for a review). Although there have been advances in the interface between motivation and learning, this work has been driven mostly by experimental techniques that explore the operation of motivational systems indirectly, rather than by computational models (Markman, Maddox and Baldwin, 2005).

This paper presents a computational model for perceptual classification that incorporates motivation. The model is based on simple perceptron learning (Rosenblatt, 1958) with modifications that instantiate the influence of motivation.

The paper first describes theories of motivation, and then presents the details of the computational model. Followed by presentation of results from model simulations and evaluation of how accurately the model fits to human data. Next, a new phenomenon predicted by the model has been demonstrated and tested against the human data, followed by analyses of model parameters. Finally, future directions for research are outlined.

Background

This section reviews *Regulatory Focus Theory* (Higgins, 2000), and a framework developed for investigating the influence of motivation on behavior (Maddox, et al. 2005).

Regulatory Focus Theory

The motivation literature distinguishes between approach and avoidance goals (Carver & Scheier, 1998; Lewin, 1935; Markman & Brendl, 2000). Goals with positive states that one wishes to achieve are called *approach goals*, while goals with negative states that one wishes to avoid are called *avoidance goals*. Higgins (1987, 1997) extended this idea by proposing *regulatory focus theory*, which suggests that - orthogonal to approach and avoidance goals - there are psychological states of readiness or sensitivity for potential gains/non-gains or losses/non-losses that tune the sensitivity of the motivational system. On this theory, a *promotion focus* is a state that focuses the motivational system on the presence or absence of gains in the environment, and a *prevention focus* is a state that focuses the motivational system on the presence or absence of losses.

Higgins and colleagues (Higgins, 1987, 1997; Higgins, et al. 1994) outlined three aspects of regulatory focus. First, *chronic focus*, is an a priori predisposition toward a promotion or prevention focus. Higgins (1987) suggested that a promotion focus is associated with a person's desire to achieve ideal states (e.g., hopes, desires, or aspirations) while a prevention focus is associated with a person's desire to achieve ought states (e.g., duties, obligations etc.). Thus, chronic focus could also describe a person's attributes.

Second, there is *situational focus* (also referred to as incentive focus), which is induced by properties of current circumstances. Someone pursuing a potential gain, is placed in a state of readiness for gain and non-gain situations in general. Likewise, someone attempting to avoid a loss is placed in a state of readiness for loss and non-loss situations in general. Situational focus can be manipulated experimentally (e.g., Crowe and Higgins, 1997; Markman, Kim, & Brendl, 2005).

A third key aspect of regulatory focus is the idea of *regulatory fit* (Avnet & Higgins, 2003; Higgins, 2000; Higgins, et al. 2003; Shah et al, 1998). A regulatory fit can occur in a number of ways. This paper concentrates on a match between the *current regulatory focus and the reward structure of the task* (also referred to as goal attainment means; Crowe & Higgins, 1997; Higgins, et al. 1994; Shah, et al. 1998). For example, in many learning experiments, people receive points for some responses and lose points for others. One might try to maximize the number of points obtained by focusing on gaining points or alternatively by focusing on avoiding the loss of points. Situations in which one can gain points involve a gain/non-gain reward structure. Situations in which one can lose points involve a loss/non-loss reward structure.

The interaction between chronic focus, situational focus, and the reward structure of the task has been studied extensively in the literature (Higgins & Spiegel, 2004). Subsets of these motivational factors have been found to affect performance in a number of tasks including problem solving anagrams (Shah et al, 1998), decision making recognition memory (Crowe & Higgins, 1997), and consumer behavior (Markman, Kim, & Brendl, 2005). The next section explains the motivational framework and the study of classification learning developed by Maddox, et al. (2005) on which the computational model was based.

Results on Classification Learning

This section describes the research on motivation and learning done by Markman, Baldwin & Maddox (2005) to explore the effects of regulatory fit on people's ability to acquire new categories.

The task used in this study was a simple one-dimensional classification task. Each stimulus is a dot that appears in a particular location on a computer screen. The categories had overlapping distributions, as shown in Figure 1. The bold solid line indicates the decision rule that yields optimal accuracy. To bias the reward structure of the task, one of the categories (category A) had a higher payoff than the other, but the payoffs for incorrect responses were the same for both categories. Thus the decision rule that optimizes reward is shifted away from the center of the high payoff category (see Figure 1). Furthermore, the categories were made difficult to learn (signal detection discriminability of the categories, d', was 1) in order to clearly show the effects of motivational manipulations on learning.

Three payoff matrices were devised to explore the influence of regulatory fit on category learning through the associated payoff matrix (Table 1). In all cases, correct responses yielded a higher payoff (or lower punishment) for one category than the other, while incorrect responses were treated equally for both categories. In the *mixed matrix*, subjects were rewarded with points for correct responses and penalized for incorrect responses. In the *gain matrix*, subjects received points for both correct and incorrect responses, though they received more points for a correct response than for an incorrect response. In the *loss matrix*, subjects lost points for both correct and incorrect responses, though they lost fewer points for correct responses than for incorrect responses.

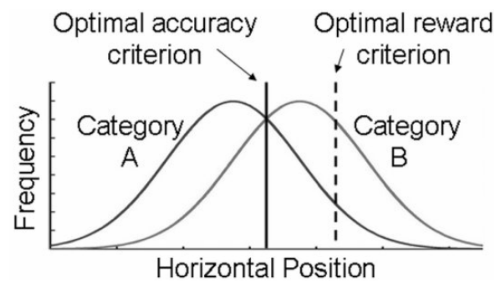

Figure 1: Category distributions and optimal decision criteria. The two categories represented here are described by one relevant dimension (position of a dot on a computer screen). Category A is associated with a higher payoff, or lower punishment, than category B. Consequently, maximizing reward requires selecting a decision criterion to the right of the criterion that maximizes accuracy.

These payoff matrices were all designed to have a signal detection decision criterion, β, of 3. Thus, the optimal classifier would use the same decision criterion across matrices. For this task the stimuli were dots presented on a computer screen, and the two categories were defined by location along an arbitrary dimension of 650 pixels. The high-payoff category had a mean of 275 and a standard deviation of 100, and the low-payoff category had a mean of 375 and a standard deviation of 100. Thus the optimal accuracy criterion was at 325 pixels and the optimal reward criterion was at 434.5 pixels.

Subjects performed three category-learning tasks, one with each of the matrices in Table 1. The regulatory focus was manipulated using a situational manipulation derived from previous experiments by Higgins and his colleagues (Higgins, 1997). In the promotion-focus condition, participants were told that in each block, they would receive an entry into a drawing to win $50 if their performance exceeded some criterion. In the prevention-focus condition, an entry ticket for a $50 raffle was displayed on the computer screen at the start of each block, and participants were told that they could keep the ticket unless their performance fell below a criterion, in which case they would lose that ticket. Thus the promotion-focus condition framed the goal as an approach state, and the prevention-focus condition framed the goal as an avoidance state. The regulatory-fit view predicts that people's performance will be closest to optimal when their regulatory focus matches the structure of the payoff matrix.

Table 1: Payoff Matrices and Performance Criteria for the Three Payoff Conditions.

Matrix	High-payoff Category		Low-payoff Category		Performance criterion (threshold)
	Correct response	Incorrect response	Correct response	Incorrect response	
Mixed	200	-100	0	-100	3,700
Gain	400	100	200	100	33,700
Loss	-111	-411	-311	-411	-43,000

Note: Subjects started each task with 0 points

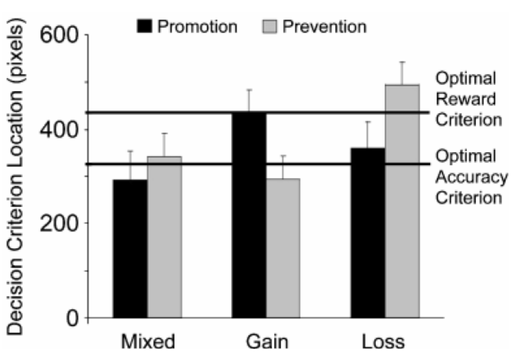

Figure 2: Mean decision criteria of promotion- and prevention-focus subjects as a function of the payoff matrix. The data were analyzed by fitting a decision-criterion model to the data from each subject in each block using 100 trials out of 150 (first and last 25 trials were discarded as noise).

The data were consistent with the regulatory fit hypothesis. Subject's performance was closest to optimal when their regulatory focus fit the payoff structure of the learning task. Subject with promotion focus had a decision criterion closer to optimal than did people with prevention focus when the payoff structure consisted of all gains. People with a prevention focus had a decision criterion closer to optimal than did people with a promotion focus when the payoff structure consisted of all losses. Performance in the two regulatory-focus conditions was roughly equivalent when the payoff structure had both gains and losses.

Based on this study the computational model was developed, as explained in the next section.

Computational Model

The building block of the computational model was perceptron (Rosenblatt, 1958) as shown in Figure 3. The perceptron is the simplest kind of feed-forward artificial neural network: a linear classifier.

A single perceptron performs binary classification, mapping its input **x** (a vector of type Real) to an output value $\hat{y}$ (a scalar of type Real) calculated as:

$$\hat{y} = < \mathbf{w}, \mathbf{x} > + b \,, \qquad (1)$$

where **w** is a vector of real-valued weights for each input, **<.,.>** denotes dot product and b is the 'bias' – a constant term.

For binary classification, the sign of $\hat{y}$ determines the class. The bias can be thought of as offsetting the activation function, or giving the output neuron a 'base' level of activity. Spatially, the bias alters the position (though not the orientation) of the decision boundary.

Learning in the perceptron is done by updating the weights with,

$$\Delta w_i = \mu(t_i - \hat{y}_i)x_i \,, \qquad (2)$$

where w_i denotes weight of the ith input x_i, μ is the learning rate and t_i is the desired output. Thus, learning involves adapting the weight vector after each iteration. The weights change only if the output $\hat{y}_i$ is different from the desired output t_i.

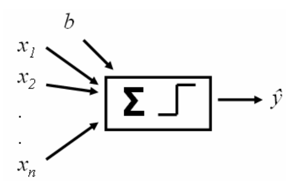

Figure 3: A single perceptron. The variables x_i are the inputs to this neuron, b is the bias and $\hat{y}$ is the output of this neuron.

The single perceptron unit can be trained accordingly for any set of two linearly separable classes (Rosenblatt, 1958). Given the classification task defined above, this perceptron predicts the decision bound to be at the optimal accuracy criterion. However, to account for the regulatory fit the simple perceptron was modified by adding an extra term to the equation that updates the weight of each input, as shown below

$$\Delta w_i = w_{acc}\mu(t_i - \hat{y}_i)x_i + w_{rew}\alpha r_i\left(m + \frac{p_o g + l p_e}{1 + |t_i - \hat{y}_i|}\right) \qquad (3)$$

where, w_{acc} is the weight on accuracy, w_{rew} is the weight on reward, μ is the learning rate for accuracy, α is the learning rate for reward, p_o and p_e depict the parameters for situational focus (promotion and prevention), g, l, and m depict payoff matrices (gains, losses, and mixed respectively) and r_i represents the reward that was attained for previous decision. The value of p_o, p_e g, and l is 1 or -1 depending upon the subject's focus and pay-off matrix. However, the value of m is either zero or one, depicting whether the subject is in or not in the mixed condition.

As represented in Eq. 3, the learning rule now consists of two parts: accuracy and reward. The accuracy part is the same as in Eq. 2 of perceptron learning, except for the new weight w_{acc} which accounts for accuracy criterion. However, the reward part needs to be discussed in more detail. Let us understand it with an example. Consider a subject given a situational promotion focus who is doing the task with a gain matrix. The parameter values for this person would be p_o = 1, g = 1, l = -1, p_e = -1, m = 0 and the Eq. 3 would reduce to

$$\Delta w_i = w_{acc}\mu(t_i - \hat{y}_i)x_i + w_{rew}\alpha r_i \frac{2}{1+|t_i-\hat{y}_i|} \quad (4)$$

The change in weight in this case increases at a higher rate in the direction of the optimal reward criterion compared to the optimal accuracy criterion. A similar pattern arises for a person with a prevention focus doing the task with a loss matrix. In contrast, for a subject in a mismatch, i.e. a subject in a prevention focus and doing a task with gain matrix or in a promotion focus and doing a task with loss matrix, Eq 3 reduces to

$$\Delta w_i = w_{acc}\mu(t_i - \hat{y}_i)x_i - w_{rew}\alpha r_i \frac{2}{1+|t_i-\hat{y}_i|} \quad (5)$$

Because of the negative sign between accuracy and reward portion, in Eq 5, the change in weight is more in the direction of the optimal accuracy criterion than the optimal reward criterion in a mismatch situation.

Further, in case of mixed pay-off matrix the factor (p_og+lp_e) would sum up to zero, independent of the situational focus, since both g and l would be 1. So in mixed condition the model would be tuned somewhere between optimal accuracy and optimal reward criterion, similar to what has been shown with the human subjects (Figure 2).

To model the data on a trial by trial basis, the weights, w_{acc} and w_{rew}, must adapt after each trial during the duration of task. This dynamic adaptation takes place according to

$$\Delta w_{acc} = \alpha(|t_i - \hat{y}_i| - (p_o g + lp_e) + \theta r_i) \quad (6)$$

$$\Delta w_{rew} = \alpha((p_o g + lp_e) - \theta r_i) \quad (7)$$

where, θ represents prediction of rewards (explained in the next paragraph). Now, in case of a regulatory fit, i.e., subject with a promotion focus who is doing a task with gain matrix or subject with a prevention focus and doing a task with loss matrix, the change in w_{acc} would be in the negative direction and change in w_{rew} would be in the positive direction (Eq 6 and 7) and vice versa for the regulatory mismatch condition. Thus a subject in a fit condition would go towards the optimal reward criterion while one with a mismatch condition would go towards accuracy.

The parameter θ is particularly interesting. In the original empirical study each subject was shown a reward meter that displayed the current number of points achieved and the distance to the reward threshold. To mimic this aspect of the study, the θ variable was introduced in the computational model to represent a prediction that the reward would be obtained. It is defined as

$$\theta = +1 \quad if \quad \frac{r_t - r_p}{q} \le r_a,$$

$$\theta = -1 \quad if \quad \frac{r_t - r_p}{q} > r_a, \quad (8)$$

where, r_t is the reward threshold (a constant defined in the beginning of the task). r_p is the reward value at any present moment, r_a is the assumed average reward, and q is the number of trials left. Thus, if the reward threshold is achievable (i.e., the ratio of difference between the reward threshold and the present reward and the number of trials left is lower than the assumed average reward) then θ = 1, and thus w_{acc} increases and w_{rew} decreases, thereby moving towards the optimal accuracy criterion. However, if the reward threshold does not seem to be achievable (i.e., the above ratio is higher than the assumed average reward) then θ = -1, and thus w_{acc} decreases and w_{rew} increases, thereby moving towards the optimal reward criterion. Thus, the parameters θ and r_a drive the model dynamically by predicting rewards and making the model chose respective action (classification in this case). This dynamical nature of the model is very similar to the reinforcement learning behavior proposed by Sutton and Barto, 1998.

Thus, the modified model has the potential to replicate and account for subject wise patterns of results shown by Markman et al. (2005).

Evaluation

The input to the model was similar to that used in the study by Markman, et al. (2005). The modified perceptron had only one input: the location of pixel, with same means and optimal reward and accuracy criterion as in the original study. The initial weight of this input was set randomly and the initial bias was random. However, these weights adapted during the trial as defined by Eqs. 3 – 8.

A total of 34 subjects out of 36 were modeled individually (17 in prevention and 17 in promotion focus), by setting up perceptrons with different initial weights. The model was not able to fit 2 subjects in prevention focus. The ordering of input was yoked to that given to the subject by Markman et al. (2005). The parameters p_o, p_e g,l, m and r_i were initialized consistently with the condition of the study in which the subject was run. Values for the other model parameters, like w_{acc} (initial value only), w_{rew} (initial value only), r_a, μ, and α, were calculated by maximizing the likelihood of the model to each subject's final reward value using a grid search of the parameter space. As shown in Table 2, the model subjects matched each human subjects final reward value accurately.

Analyses of these results lead to two interesting observations. First, the model shifted strategies over the course of the study in a systematic way. When there was a regulatory fit (promotion/Gain or prevention/Loss) and equation 8 indicated that the reward threshold was likely to be reached, the model often shifted from a focus on the reward criterion to a focus on the accuracy criterion.

Similarly in a mismatch condition (promotion/Loss and prevention/Gain), if equation 8 suggested that the reward threshold was unlikely to be reached, the model often shifted from an optimal accuracy criterion to an optimal reward criterion. Motivated by this observation, the original human subject data by Markman et al. (2005) was reanalyzed and these shifts were indeed found in the data (Table 3). This novel prediction would help in designing new studies, which in turn could give an insight on the dynamics of motivational influences on learning.

Table 2. Standardized error in predicting final reward values for each human subject.

Regulatory Focus/Pay-Off Matrix	Standardized Error
Prevention/Gain	0.145
Prevention/Loss	0.177
Prevention/Mixed	0.112
Promotion/Gain	0.225
Promotion/Loss	0.277
Promotion/Mixed	0.167

Table 3. Percentage accuracy of the model in predicting shifts by human subjects, as well as the number of subjects that shifted in each condition.

Regulatory Focus/Pay-Off Matrix	Percentage Accuracy (in predicting shifts)	Number of Subjects Who Shifted
Prevention/Gain	70.58	9
Prevention/Loss	70.58	11
Prevention/Mixed	64.70	11
Promotion/Gain	64.70	6
Promotion/Loss	64.70	8
Promotion/Mixed	64.70	7

Second, the parameter values of the model at various times during the run provide interesting insights in to the motivation-learning interface. For example, consider the parameter w_{acc}. The initial value for w_{acc} was approximately the same in all four conditions. However, at the end of the trial, it was significantly higher in the mismatch conditions as compared to the match conditions. A two-way ANOVA on the weight parameter values by condition revealed a significant interaction ($F(1,34)=7.722$, $p<0.05$). As shown in Figure 3, the interaction in the model reflects the interaction found by Markman et al. (2005), i.e. people in regulatory fit tend to move their decision criterion towards the optimal reward criterion.

Additionally, the value of r_a, which is used to predict the future reward and achievability of the task, describes an important characteristic of human behavior: reward/risk analysis. If the value of r_a for a given subject was higher than the average value of rewards (based on payoff matrix), then that subject can be said to be careless. Conversely, if the r_a value was lower than the average then the subject was cautious. Figure 4 shows the mean r_a values for the six different combinations. A two-way ANOVA on r_a values by payoff matrix revealed a significant interaction ($F(1,34)=70.3$, $p<0.05$). It is clear from the figure that the mean r_a values for the loss payoff matrix are lower than the average rewards, thereby indicating that when people do tasks with loss payoff matrix they tend to become cautious. In contrast, in the gain payoff matrix the mean r_a values are higher than the average rewards, and subjects are less cautious. This behavior could also provide a reason for the variation (depending upon the condition) in shifting of strategies during the task. Subjects who follow a cautious approach (or loss payoff matrix) should shift more in the end compared to subjects with a less cautious approach (or gain payoff matrix). Therefore, just based on the r_a parameter of the model, it is possible to predict that if a person is in a prevention or promotion regulatory focus, the chances of shifting are greater in the case of a loss payoff matrix compared to a gain payoff matrix. This prediction coincides with the human data (Table 3).

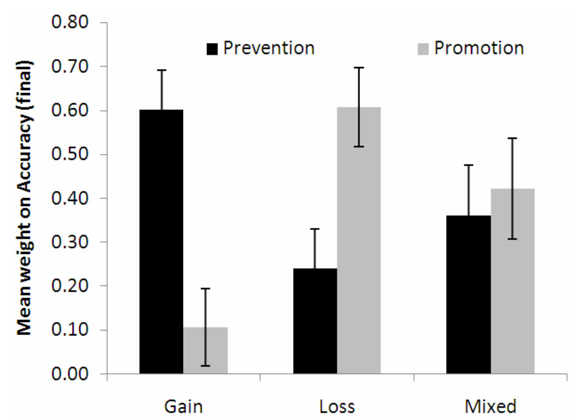

Figure 3. Mean weight on accuracy (w_{acc}) for different conditions at the end of trial. The interaction in the model reflects the interaction found by Markman et al. (2005), i.e. people in regulatory fit tend to move their decision criterion towards the optimal reward criterion.

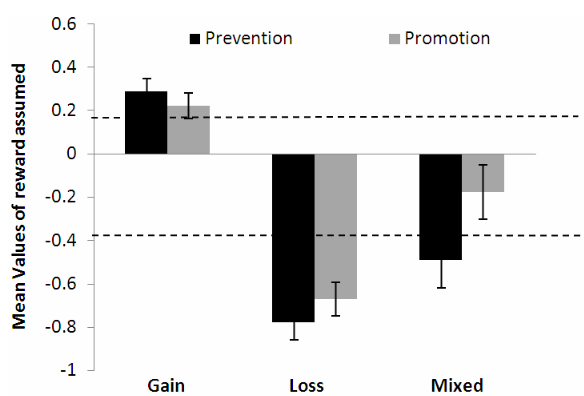

Figure 4. Mean reward assumption (r_a) for different conditions. The dashed lines indicate average reward values (positive for gains and negative for losses).

Finally, the learning rate for reward, α, was not significantly different in the four conditions. However the learning rate for accuracy, μ, was significantly higher (two-way ANOVA by condition, $F(1,34)=6.22$, $p<0.05$) for people in a regulatory fit than in a mismatch. When the subjects came to do the classification task they only knew that can perform well if they accurately classify the instances. The manipulation of regulatory fit was completely unknown to them. The high learning rate for accuracy shows that the people in regulatory fit are more flexible than the ones in regulatory mismatch, and that they are more prone to shifting back to accuracy in the end. Figure 5 shows the mean learning rates for accuracy in different conditions.

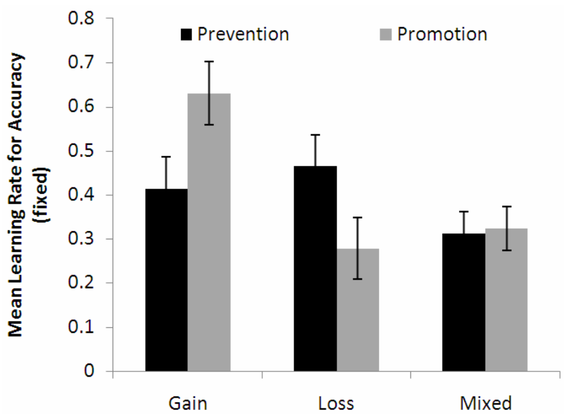

Figure 5. Mean learning rate of accuracy criterion (μ) for different conditions. High learning rate for accuracy shows that the people in regulatory fit are more flexible than the ones in regulatory mismatch.

Conclusion

The computational model presented in this paper elaborates the motivation-learning interface observed by Markman et al. (2005). To our knowledge, this model is the first to incorporate the effects of regulatory focus on learning. The model accurately fits the individual subject data, and provides a detailed insight into the motivation-learning interface. The model led to the discovery of an unstudied phenomenon of shifting strategies across trials. These predictions were supported by a re-analysis of the original human subject data, thus elaborating and strengthening the theory proposed by Maddox et al. (2005). Furthermore, the evaluation of various parameters of the model not only elaborates the underlying phenomenon but also suggests a number of important avenues for future research that we plan to pursue. These results constitute a first computational step towards the understanding how motivational influences on learning and cognition.

Acknowledgments

This research was supported by AFOSR grant FA9550-06-1-0204 to WTM and ABM, an IC2 Institute fellowship to ABM, and NSF grant EIA-0303609 to RM.

References

Avnet, T., & Higgins, E. T. (2003). Locomotion, assessment, and regulatory fit: Value transfer from "how" to "what." *Journal of Experimental Social Psychology*.

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

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes*.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*.

Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of Personality & Social Psychology.*

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist.*

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

Higgins, E. T., Idson, L.C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology.*

Higgins, E. T., & Spiegel, S. (2004). Promotion and prevention strategies for self-regulation: A motivated cognition perspective. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self regulation: Research, theory and applications*. New York: Guilford Press.

Lewin, K. (1935). A dynamic theory of personality. *New York: McGraw-Hill.*

Maddox, W. T., Baldwin, G. C., & Markman, A. B. (2005). Regulatory focus effects on cognitive flexibility in rule-based classification learning. *Memory and Cognition*.

Maddox, W.T., Markman, A.B., & Baldwin, G.C. (2006). Using classification to understand the motivation-learning interface. *Psychology of Learning and Motivation*.

Markman, A. B., & Brendl, C. M. (2000). The influence of goals on value and choice. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation*. San Diego, CA: Academic Press.

Markman, A. B., Kim, K., & Brendl, C. M. (2005). The influence of goal activation on preference. *Manuscript in preparation*.

Markman, A. B., Baldwin, G. C., & Maddox, W. T. (2005). The interaction of payoff structure and regulatory focus in classification. *Psychological Science.*

Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2005). The implications of advances in research on motivation for cognitive models. *Journal of Experimental and Theoretical Artificial Intelligence.*

Rosenblatt, F.(1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review.*

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.*

Sutton, Richard S.; Andrew G. Barto (1998). Reinforcement Learning: An Introduction. MIT Press. ISBN 0-262-19398-1.