# Emotion-Based Reinforcement Learning

**Woo-Young Ahn**[1] **(ahnw@indiana.edu)**
**Olga Rass**[1] **(rasso@indiana.edu)**
**Yong-Wook Shin**[2] **(shaman@amc.seoul.kr)**
**Jerome R. Busemeyer**[1] **(jbusemey@indiana.edu)**
**Joshua W. Brown**[1] **(jwmbrown@indiana.edu)**
**Brian F. O'Donnell**[1] **(bodonnel@indiana.edu)**
[1]Department of Psychological and Brain Sciences, Indiana University
[2]Department of Psychiatry, Ulsan University School of Medicine

## Abstract

Studies have shown that counterfactual reasoning can shape human decisions. However, there is a gap in the literature between counterfactual choices in description-based and experience-based paradigms. While studies using description-based paradigms suggest participants maximize expected subjective emotion, studies using experience-based paradigms assume that participants learn the values of options and select what maximizes expected utility. In this study, we used computational modeling to test 1) whether participants make emotion-based decisions in experience-based paradigms, and 2) whether the impact of regret depends on its degree of unexpectedness as suggested by the current regret theory. The results suggest that 1) participants make emotion-based choices even in experience-based paradigms, and 2) the impact of regret is greater when it is expected than when it is unexpected. These results challenge the current theory of regret and suggest that reinforcement learning models may need to use counterfactual value functions when full information is provided.

**Keywords:** Decision making; Bayesian modeling; mathematical modeling; regret; reinforcement learning.

## Introduction

In our daily lives, we constantly face decisions to make and assess the costs and benefits of possible options (e.g., "Should I buy a lottery or just buy a snack with this money?", "Should I buy Apple or Google stock?"). Usually we know only the outcome of our choices. On rare occasions, we also know what would have happened if we had made different choices (e.g., stock market). Having 'complete feedback' (or *full information*) under risk or uncertainty can evoke strong emotions such as regret or disappointment that are triggered by our capacity to reason counterfactually.

The effects of counterfactual reasoning have received much attention, and several theories have been proposed. A growing consensus suggests that disappointment and elation are elicited by comparison between different states (e.g., "my grant was not funded...") whereas regret and rejoice come from comparison between different choices (e.g., "I should have married another person..."). Also, the unique aspect of regret is a feeling of *responsibility* that comes with negative outcomes from choices.

Among several theories of counterfactual decision-making, *decision affect theory* is regarded as one of the leading models (Mellers, Schwartz, & Ritov, 1999). Decision affect theory assumes that individuals make emotion-based choices and want to maximize subjective expected *pleasure* (or emotion) rather than to maximize expected return. In decision affect theory, our emotional responses ($R$) are based on obtained outcomes, relevant comparisons, and beliefs about the likelihood of the outcomes:

$$\begin{aligned} R \quad \propto \quad & \text{Chosen Outcome Utility} + \\ & \text{Regret / Rejoice} + \\ & \text{Disappointment / Elation} \end{aligned} \tag{1}$$

All counterfactual terms (regret, rejoice, disappointment, and elation) are weighted by their unexpectedness. Decision affect theory effectively explained various experimental results (Mellers et al., 1999) and Coricelli et al. (2005) used a modified version of the theory to examine the neural correlates of regret using description-based paradigms.[1]

Several studies have examined counterfactual decision-making using experience-based paradigms as well (Lohrenz, McCabe, Camerer, & Montague, 2007; Boorman, Behrens, & Rushworth, 2011; Hayden, Pearson, & Platt, 2009; Yechiam & Rakow, 2011). Although models used in the studies differ slightly from each other, all previous studies used reinforcement learning models, which assume that participants learn about chosen and foregone outcomes from trial-by-trial experience and then choose an option that has the highest expected value.

This study was developed from this gap in the literature: to explain choice behaviors in description-based paradigms with full information, researchers have assumed participants would make emotion-based choices. To explain choice behaviors in experience-based paradigms, researchers have assumed that participants learn the obtained and foregone payoffs and do not make emotion-based choices. We tested whether individuals make emotion-based choices in experience-based paradigms by building computational models for all competing hypotheses. This approach allowed us to quantitatively compare hypotheses in a rigorous way.

Another aim of the study was to test whether regret would be weighted by its unexpectedness (i.e., surprisingness). Mellers et al. (1999) claimed that "...unexpected out-

[1]In description-based paradigms, the outcomes of all options and their probabilities are provided to participants and participants rarely receive feedback. In experience-based paradigms, participants must learn the outcomes or their probabilities from their personal experience (Hertwig, Barren, Weber, & Erev, 2004).

comes have greater emotional impact than expected outcomes." However, how would you feel given the following scenarios? In scenario 1, an Apple employee told you some inside information about Apple, which would increase its stock price. You believed that this was 80% reliable, but you did not buy the stock whose price sky-rocketed. In scenario 2, an untrustworthy looking stranger told you the same information. You believed he was 20% reliabilible, but you did not buy the stock, whose price sky-rocketed. According to Mellers et al. (1999), you would experience more regret in scenario 2. However, we hypothesized that scenario 1 would generate more regret because of the unique aspect of regret: a feeling of responsibility. Therefore, we predicted that regret would be weighted by its *expectedness* rather than its unexpectedness. Mellers, Schwartz, Ho, and Ritov (1997) showed that a smaller probabilities of disappointment/elation were associated with greater emotional response. Although Mellers et al. (1999) claimed that the effect of probability would be the same with regret/rejoice, no experiment has directly tested it to our knowledge.

In sum, we designed our experiment to test the following hypotheses. The first hypothesis proposes that participants will learn the chosen and fictive outcomes, compare all available options, and try to maximize their expected return ("Fictive Learning Alone"). The second hypothesis proposes that participants will make emotion-based decisions (i.e., maximize their expected subjective emotion) and their regret will be weighted by its unexpectedness ("Original Regret"). The third hypothesis proposes that participants will make emotion-based decisions and will weight their regret by its expectedness ("Modified Regret"). We designed our experiment to test these hypotheses.

# Method

## Participants

Nineteen healthy individuals (7 men, mean age = 23.0, SD=4.9) participated in the study. Electroencephalography (EEG) was continuously recorded from the scalp, but EEG findings are not reported in this paper. Participants were paid $10/hr for participation and told that they would earn performance bonuses based on total points earned during the task. In reality, all participants received a fixed amount ($5) as their bonus money (Lejuez et al., 2003). Study procedures were approved by the Indiana University's Human Subjects Institutional Review Board.

## Task

All participants completed four separate gambling games, the order of which was randomly mixed for each participant. At the start of each game, participants were told that each game was independent of the previous game(s). In each game (90 trials/game), participants were asked to choose one of two options. One option was a safe option in which participants always won a fixed amount of points (e.g., 11). The other was a risky option in which participants won either larger (e.g., 26) or smaller points (e.g., 1). The probability of winning larger points was fixed but unknown, and had to be learned from experience. The payoffs of both chosen and unchosen options were revealed on every trial ("full information"). The locations of the options were fixed within games, but randomized across games. Participants were encouraged to choose an option that would maximize their gain. Payoffs were distributed so that the long-term expected values of two options were the same (see Table 1).

Table 1: The payoff distributions of games 1-4. Note that the (long-term) expected values of the safe option (M) and the risky option are the same. M: points of the safe option, L: low (smaller) points, H: high (larger) points, %H: the probability of winning larger points. SD: standard deviation.

| Game | M | Risky Option | | | | |
|---|---|---|---|---|---|---|
| | | L | H | %H | Mean | SD |
| 1 | 12 | 1 | 56 | 0.2 | 12 | 22.0 |
| 2 | 11 | 1 | 26 | 0.4 | 11 | 12.3 |
| 3 | 10 | 1 | 16 | 0.6 | 10 | 7.4 |
| 4 | 9 | 1 | 11 | 0.8 | 9 | 4.0 |

The timing and presentation of a trial is illustrated in Figure 1. Each trial started with a message ("WAIT"), which was was presented for 1-1.5s. After two options were presented, the participant had 2s to select an option by pressing buttons corresponding in a spatially compatible way to the options. The color of the chosen option remained changed for .6s, and the payoffs of both options appeared for 1s.

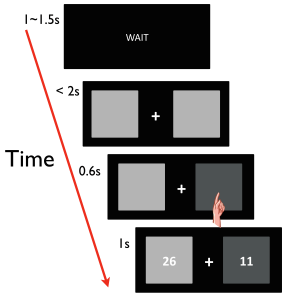


Figure 1: Time course of the gambling task.

## Computational Modeling

Three hypotheses (1. Fictive Learning Alone, 2. Original Regret model, 3. Modified Regret model) were implemented as three distinct reinforcement learning models. They utilized identical learning (probability learning) and choice rules (softmax), but used different value functions. Due to the specific design of the task (only 2 possible payoffs of the risky option in each game), it was assumed that participants would learn the probability of a larger payoff of the risky option (probability learning). In the delta rule (Rescorla & Wagner,

1972), the probability of a larger payoff (H) (the risky option) on the next trial $t+1$, $Pr_H(t+1)$, is updated as follows:

$$Pr_H(t+1) = Pr_H(t) + \gamma \cdot [Y(t) - Pr_H(t)] \quad (2)$$

Here $\gamma$ is the learning rate ($0 < \gamma < 1$) and $Y(t)$ is the outcome (1 if H, 0 if L) of the current trial $t$. We assumed no learning occurred about the safe option because its payoff was always the same (e.g., 11) in a given game. We assumed that the choice of a risky or safe option did not affect the learning rate.[2]

Action selection was implemented via the Luce choice rule (a.k.a. softmax) (Luce, 1959). The inverse temperature parameter ($\theta$) determines the sensitivity of the choice probabilities to the action values. We employed a trial-independent choice rule (Yechiam & Ert, 2007), where $\theta = 3^c - 1$ ($0 < c < 5$). When $c$ approaches zero, choices become completely random (exploratory). When $c$ becomes large, choices become deterministic (exploitive).

$$Pr_R(t+1) = \frac{e^{\theta \cdot Q_R(t+1)}}{e^{\theta \cdot Q_R(t+1)} + e^{\theta \cdot Q_S(t+1)}} \quad (3)$$

Here $Q_R(t+1)$ and $Q_S(t+1)$ are action values of choosing the risky (R) and safe (S) options on trial $t+1$, respectively. $Pr_R(t+1)$ is the probability of choosing the risky option on trial $t+1$. Next, we describe differences between three competing models (1. Fictive Learning Alone (FLA), 2. Original Regret model, 3. Modified Regret model).

**Fictive Learning Alone (FLA)** The FLA model assumes that participants compute action values of each option separately, then select an option that would maximize their expected return. The action value for the safe option is always the same on each game, $Q_S(t+1) = M^\alpha$ ($0 < \alpha < 1.5$). In other words, the chosen outcome utility of $X$ points ($u_X$) was set to $X^\alpha$. $\alpha$ is a parameter that governs the shape of the utility function. As $\alpha$ goes to zero, the reward sensitivity diminishes. The action value of the risky option is the sum of two possible utilities, weighted by their probabilities. In other words, $Q_R(t+1) = u_H \cdot Pr_H(t+1) + u_L \cdot Pr_L(t+1)$.[3] These action values are entered into Equation 3 to compute the probability of choosing each action on the next trial.

**Original Regret Model** In Regret models (both Original and Modified versions), it is assumed that participants choose an option that maximizes their *subjective expected pleasure* or *emotion* (Mellers et al., 1999). Thus, action values are the weighted sum of expected *emotional responses* ($R$ in Equation 1), rather than expected *utilities*.

Here we used the notation that $R_{A(B)}(t+1)$ is the expected emotional response on trial $(t+1)$ when chosen and unchosen payoffs are $A$ and $B$, respectively. We used Equation 1 to calculate $R_{M(L)}(t+1)$, $R_{M(H)}(t+1)$, $R_{L(M)}(t+1)$, and $R_{H(M)}(t+1)$.[4] Following Mellers et al. (1999), we set regret/rejoice and disappointment/elation terms to $sgn(A-B) \cdot |A-B|^\alpha$ when chosen and unchosen payoffs were $A$ and $B$.[5] We assumed that $\alpha$ is identical for both counterfactual functions and the chosen outcome utility. Importantly, regret/rejoice or disappointment/elation will be weighted by its surprisingness. We used 1 minus its probability as an index of surprisingness (e.g., 1-$Pr_H(t+1)$) (Mellers et al., 1999). For example, suppose a participant chooses the safe option (chosen payoff=$M$ and the foregone payoff = $H$). Then, the expected emotional response can be expressed as $R_{M(H)}(t+1)$ from Equation 1, which is equal to $M^\alpha + (-1) \cdot |M-H|^\alpha \cdot (1 - Pr_H(t+1))$.[6] If the participant chooses the risky option and the chosen payoff is $L$, the expected emotional response is $R_{L(M)}(t+1)$. $R_{L(M)}(t+1)$ is equal to $L^\alpha + (-1) \cdot |L-M|^\alpha \cdot (1 - Pr_L(t+1)) + (-1) \cdot |L-H|^\alpha \cdot (1 - Pr_L(t+1))$. Note that the disappointment term was included in this case. $R_{M(L)}(t+1)$ and $R_{H(M)}(t+1)$ can be calculated in the same way and these terms can be used to calculate action values in Equation 4:

$$\begin{aligned} Q_S(t+1) &= R_{M(H)}(t+1) \cdot Pr_H(t+1) + R_{M(L)}(t+1) \cdot Pr_L(t+1) \\ Q_R(t+1) &= R_{H(M)}(t+1) \cdot Pr_H(t+1) + R_{L(M)}(t+1) \cdot Pr_L(t+1) \end{aligned} \quad (4)$$

The computed action values are entered into the softmax choice rule in Equation 3 to calculate trial-by-trial probability of choosing a risky (or safe) option.

**Modified Regret Model** This model is identical to the Original Regret model except that regret (but not any other counterfactual comparisons) is weighted by Regret's expectedness. We used regret's probability as its expectedness (e.g., $Pr_H(t+1)$). Thus, only $R_{M(H)}(t+1)$ and $R_{L(M)}(t+1)$ are different between two Regret models because participants experience rejoice, but no regret for $R_{M(L)}(t+1)$ and $R_{H(M)}(t+1)$.

**Summary of Three Competing Models** In sum, we compared three different models (specifically, value functions). The FLA model assumes that participants evaluate two options separately and choose the option that maximizes their expected return. The two Regret models assume that participants evaluate anticipated emotional responses and maximize their subjective pleasure. The Regret models, however, make different assumptions about the role of surprisingness when processing regretful outcomes. All three models have three free parameters: learning rate ($\gamma$), utility shape ($\alpha$), and choice consistency ($c$). We used hierarchical Bayesian approach to estimate them, which is useful for reliably estimating group and individual parameters (for a review see Lee, 2011).

[2] We tried several other versions of learning rules (e.g., separate learning rates for chosen and unchosen options) and choice rules (e.g., trial-dependent inverse temperature parameter) that are not reported here, but they did not improve model-fits.

[3] $Pr_L(t+1) = 1 - Pr_H(t+1)$, $u_H = H^\alpha$, and $u_L = L^\alpha$.

[4] In all settings, $L < M < H$ (e.g., L=1, M=11, H=26).

[5] $sgn(x)$ = 1 if $x > 0$, -1 if $x < 0$.

[6] The disappointment/elation term is present only for risky choices. The disappointment/elation term is missing in $R_{M(H)}(t+1)$ because the safe option was chosen, in which there is only one state.

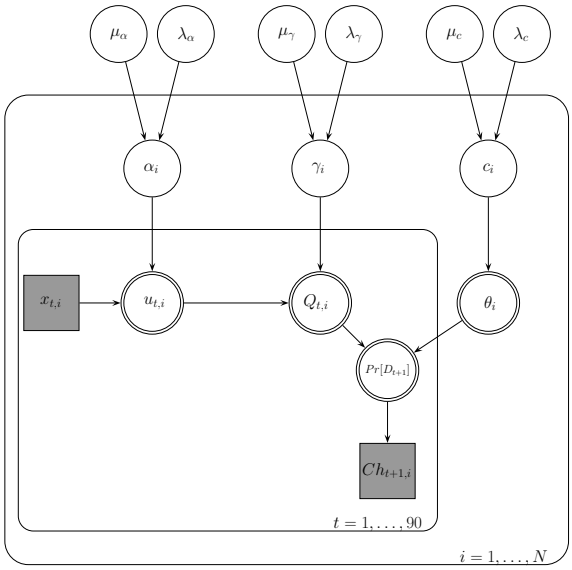


Figure 2: Graphical depiction of the hierarchical Bayesian analysis for three reinforcement learning model. $R_{A(B)t,i}$ replaces $u_{t,i}$ for Regret models.

## Graphical Model Implementation - Hierarchical Bayesian Parameter Estimation

Figure 2 shows the graphical representation of all three models. We modeled the variation in $\gamma_i$, $\alpha_i$, and $c_i$ parameters by assuming they have censored Gaussian distributions across participants. (e.g., $\gamma_i \sim Normal(\mu_\gamma, \lambda_\gamma)I(0,1)$, where $\mu_\gamma$ and $\lambda_\gamma$ are the mean and precision variables of the Gaussian distribution). Mean variables had uniform priors and precision variables had Gamma priors (e.g., $\lambda_\gamma \sim Gamma(.001, .001)$). In Figure 2, clear and shaded shapes indicate latent variables and observed variables, respectively. Single and double outlines indicate probabilistic and deterministic functions of input, respectively. Circles and squares indicate continuous and discrete variables, respectively (Lee, 2008). Vectors $x_{t,i}$ (payoffs) and $Ch_{t+1,i}$ (choices) were observed and individual ($\gamma_i$, $\alpha_i$, $c_i$) and group parameters ($\mu_\gamma$, $\mu_\alpha$, $\mu_c$, $\lambda_\gamma$, $\lambda_\alpha$, $\lambda_c$) were estimated. We used OpenBUGS (Lunn, Spiegelhalter, Thomas, & Best, 2009) to perform Bayesian inference. We used 50,000 posterior samples collected following a total of 30,000 burn-in samples. Multiple chains were used to check convergence and $\hat{R}$ values indicated that Markov chain Monte Carlo (MCMC) chains converged well with the target posterior distributions. Given that participants' choice behavior varied across games (see Figure 3), we estimated parameters separately for each game (but across all participants within each game). Ideally, model parameters should remain stable across games. Otherwise the model might simply mimic data without providing a coherent theoretical explanation of choice behavior.

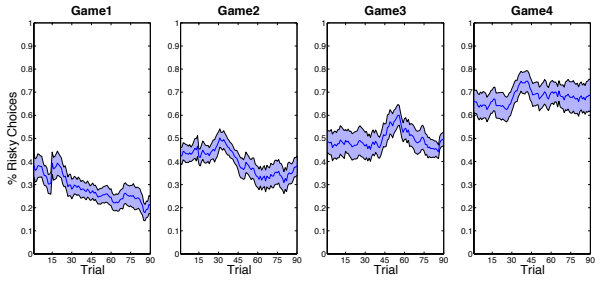


Figure 3: The mean proportions of risky choices over trials on Games 1-4. The blue solid line indicates the group mean on each trial and shaded region indicates ±s.e.m. (a moving-average filter was used).

# Results

## Behavioral Results

The proportions of risky choices in each game are plotted in Figure 3. As seen, participants' choice behavior varied across games although the expected values of two options were equated on all games. The mean proportions of risky choices on games 1-4 were .28, .40, .50, and .68 and the differences between games were all significant (games 1 vs 2: $p < .003$; games 2 vs 3: $p < .004$; games 3 vs 4: $p < .001$).

Next, we examined the effect of chosen feedback, foregone feedback, and the magnitude of their difference (Coricelli et al., 2005). For this goal, we performed panel logic regression using the individual random-effects model. The dependent variable was 'switch' (1 if switched from the previous trial, 0 otherwise), and independent variables were the chosen payoffs (or feedback) ($fb$), the foregone payoffs ($fgFb$), and the magnitude of their difference ($|fb - fgFb|$) on the previous trial (T-1). Table 2 shows that participants were more likely to switch if the chosen feedback was lower ($p <$ 3E-16), the foregone feedback was higher ($p <$ 2E-13), and the magnitude of the difference was higher ($p < .011$). These results suggest that participants take all three variables into account when making decisions.

To examine the effect of feedback on previous trials, another panel logistic regression analysis was performed, examining how many previous trials ($fb - fgFb$) biased the switch behavior. Figure 4 shows that chosen−foregone payoffs of up to two previous trials significantly influenced the switch behavior.

Table 2: Regression analysis (panel logit procedure with individual random effect). fb: the chosen payoff (feedback), fgFb: the foregone payoff (feedback).

| Variable | Coefficient | Std. Error | $t$ | $p$ |
|---|---|---|---|---|
| Constant | .3902 | .0186 | 21.00 | **<3E-16** |
| fb | -.0064 | .0009 | -7.44 | **<2E-13** |
| fgFb | .0027 | .0007 | 3.71 | **<.001** |
| \| fb -fgFb \| | .0025 | .0010 | 2.53 | **.011** |

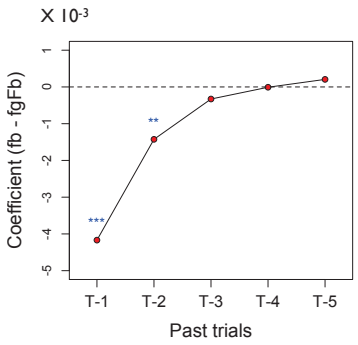


Figure 4: Effects of past outcomes on current choice behavior. fb: the chosen payoff, fgFb: the foregone payoff. *** $p<.0001$ ** $p<.001$.

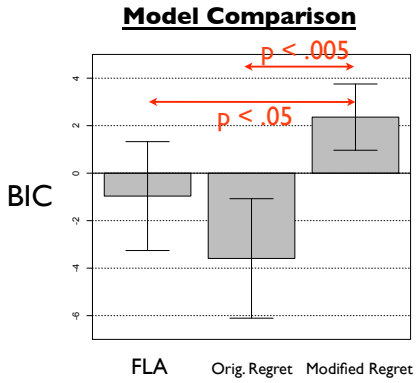


Figure 5: BIC (Bayesian information criterion) scores of three competing models compared to the baseline model. Note that higher BIC indicates a better model fit. Error bars indicate ±s.e.m. FLA: Fictive Learning Alone.

## Modeling Results

To determine which model best fits our data, we used maximum likelihood estimation (MLE) methods to fit the model to each person and game separately, and then used the Bayesian information criterion (BIC) (Schwartz, 1978) to compare the Bernoulli baseline model, in which the probabilities of two options were equal to the individual's overall proportion of each option (the number of free parameters=1) against three models of interest.[7] The BIC score is a statistic that combines badness of fit with a penalty for the number of parameters. To evaluate the models, we used a BIC change score that measures the improvement of the computational model over the Bernoulli baseline model (BIC change equals the BIC from the baseline model minus the BIC from the cognitive model). Therefore positive BIC changes represent improvement over baseline, and the model with the highest BIC change is considered the best.

Figure 5 shows that the Modified Regret model has the best model fit. When tested across participants, the difference was significant (the Modified vs. Original Regret models: $p < .005$, the Modified Regret vs. FLA models: $p < .05$). When the descriptive accuracy was assessed by posterior predictive analysis, the best-fitting model (the Modified Regret model) provided good individual-level model predictions. For example, Figure 6 illustrates a good match between the observed data (Figure 6A) and the model's predictions for a participant's choices (Figure 6B).

Next, we examined whether the parameter values of three models would remain stable across games. Again, ideally model parameters should be similar across different games or tasks. In Figure 7, all parameters of the models were plotted across games 1-4. Clearly, the parameters of the Modified Regret model, which had the best model fit, were the most stable across games. Note that the utility shape ($\alpha$) and consistency ($c$) parameters of FLA and Original Regret models varied greatly across games. In sum, the results of both model fit and parameter consistency indicate that the Modified Regret model explains participants' choice behavior best.

[7]We are currently working on comparing models by estimating their Bayes factors (Kruschke, 2011)

## Discussion

The goals of this study were to examine: (1) whether participants make emotion-based choices in experience-based paradigms; (2) whether regret would be weighted by its unexpectedness or expectedness. The modeling results provided strong support for the Modified Regret model: the model had the best model fit and its parameters were the most stable across games, suggesting it might provide a coherent theoretical account for choice behavior across games. The results provide strong support that participants make emotion-based choices and experience greater regret when it was expected rather than when it was unexpected.

We believe this study is the one of the first attempts to incorporate emotion-based decisions into reinforcement learning. Our findings are consistent with previous studies using description-based paradigms that found participants made emotion-based decisions. Our results suggest that reinforcement learning models may need to use value functions that can incorporate emotional components. The results are also consistent with the notion that emotions provide a common currency on how we make decisions under risk or uncertainty (Loewenstein, Weber, Hsee, & Welch, 2001; Weber & Johnson, 2009).

We also believe these results need to be tested in other experience-based paradigms and to determine their generalizability. Some studies found that Bayesian learning models outperformed the delta learning rule (Boorman et al., 2011). Although it is possible that using such a learning model can improve the model fit for all three models, we do not think it will change the main findings of the current study. In sum, we found strong support for the Modified Regret model, which challenges the current theory of regret.

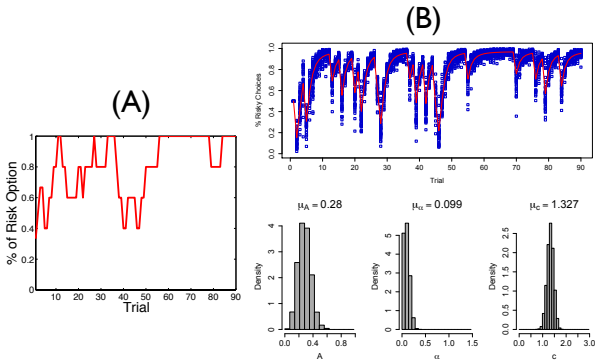


Figure 6: Posterior predictive assesment of the Modified Regret model for one participant. (A) The participant's proportion of risky choices over trials (smoothed with a moving-average filter) (B) posterior predictive distributions for $Pr_R(t)$. Small blue squares indicate 50 random samples from the posterior predictive distributions. The red solid line indicates the mean values of the distributions. The participant's model parameter values are in the bottom figure.

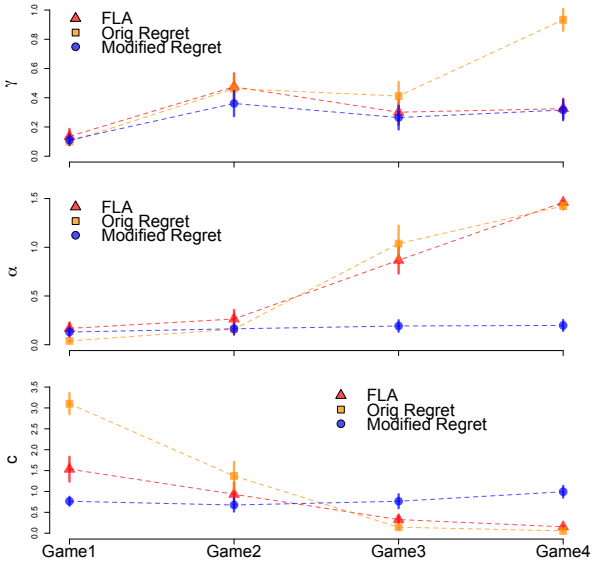


Figure 7: Parameter values of three competing models across games 1-4. Symbols and error bars indicate the means and standard deviations of the posterior distributions, respectively. FLA: the Fictive Learning Alone model.

## Acknowledgments

This work was supported by the NIMH (R01 MH62150 to BFO), NSF (0817965 to JRB), and Indiana University College of Arts and Sciences Dissertation Fellowship (to WYA).

## References

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biology*, *9*(6), e1001093.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nature Neuroscience*, *8*, 1255–1262.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science*, *324*, 948–950.

Hertwig, R., Barren, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534-539.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Academic Press / Elsevier.

Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin and Review*, *15*(1), 1.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7.

Lejuez, C., Aklin, W., Jones, H., Richards, J., Strong, D., Kahler, C., et al. (2003). The balloon analogue risk task (bart) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology*, *11*(1), 26.

Loewenstein, G., Weber, E., Hsee, C., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267.

Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. U.S.A.*, *104*, 9493–9498.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, *28*(25), 3049–3067.

Mellers, B., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory. *Psychological Science*, *8*(6), 423–429.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*(3), 332.

Rescorla, R. A., & Wagner, A. R. (1972). *A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement* (A. H. Black & W. F. Prokasy, Eds.). Appleton-Century-Crofts.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *5*, 461-464.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, *60*, 53–85.

Yechiam, E., & Ert, E. (2007). Evaluating the reliance on past choices in adaptive learning models. *Journal of Mathematical Psychology*, *51*, 75-84.

Yechiam, E., & Rakow, T. (2011). The effect of foregone outcomes on choices from experience. *Experimental Psychology*, 1–13.