Choice adaptation to increasing and decreasing event probabilities

Samuel Cheyette (sjcheyette@gmail.com)
Dynamic Decision Making Laboratory, Carnegie Mellon University
Pittsburgh, PA. 15213
Emmanouil Konstantinidis (em.konstantinidis@gmail.com)
School of Psychology, University of New South Wales
Sydney, NSW 2052, Australia
Jason L. Harman (jharman@lsu.edu)
Department of Psychology, Louisiana State University, Baton Rouge, LA. 70803
Cleotilde Gonzalez (coty@cmu.edu)
Dynamic Decision Making Laboratory, Carnegie Mellon University
Pittsburgh, PA. 15213

Abstract

A constant element of our modern environment is change. In decision-making research however, very little is known about how people make choices in dynamic environments. We report the results of an experiment where participants were asked to choose between two options: a dynamic and risky option that resulted in either a high or a low outcome, and a stationary and safe option that resulted in a medium outcome. The probability of the high outcome in the risky option decreased or increased linearly over the course of the task while the probability of the medium outcome stayed the same throughout. We find that adaptation to change is related to the direction of that change, and that the way people adapt to changing probabilities relates to their willingness to explore available options. A cognitive model based on Instance-Based Learning Theory reproduces the behavioral patterns.

Keywords: Change; Dynamic Decisions; Adaptation; Instance-Based Learning Theory; Decisions from Experience

Introduction

More than ever before the world around us seems to be changing rapidly. Technology has contributed to increasing availability of information and connectedness that contribute to a sense of rapid change. We make decisions in constantly changing situations and our ability to detect and adapt to those changes may determine the success of our choices. For example, a broker in the stock market must be sensitive to the changes in the interest rates in an attempt to maximize the long-term investments gains. In the context of reinforcement learning and restless bandit tasks, researchers have investigated change in similar settings, such as adaptation and detection of change, and exploration-exploitation tradeoffs in dynamic environments (e.g., Gureckis & Love, 2009). Yet, relatively little is known about how humans detect change when the change occurs gradually, and particularly when making decisions from experience while aiming at maximizing long-term gains.

Dynamic decision theory was first introduced by Ward Edwards (1962) who argued for the study of dynamic situations in which decision makers confront a sequence of decisions, and in which the environment changes while a decision maker is evaluating possible courses of action. Yet, to this date behavioral work on how humans behave under changing conditions and how we adapt to change is relatively limited.

About a decade ago, research started to shift from the overwhelmingly popular study of one-shot decisions to the study of repeated and consequential choice. In repeated choice conditions, early decisions produce payoffs and information that may influence future choices. This is one of the reasons that researchers have focused on experience and cognitive processes such as learning, memory, and recognition as key psychological processes of dynamic decision-making (Gonzalez, Lerch, & Lebiere, 2003).

The study of *decisions from experience* has expanded considerably in the past years, perhaps due to the development of simple experimental paradigms (e.g., "clicking paradigms") which have been used to study choice in its most essential form: in binary conditions (Baron & Erev, 2003; Hertwig et al., 2004). These paradigms involve the selection between two options, in the absence of descriptions of possible outcomes and probabilities. For example, in a repeated consequential choice paradigm (Baron & Erev, 2003), participants select between two buttons a fixed number of times (e.g., 100 times). After each selection, an outcome is displayed (i.e., feedback). This outcome is the realization of a probability distribution assigned to the button selected, which is unknown to participants. This paradigm is illustrated in Figure 1.

Using this experimental paradigm, researchers have investigated a number of issues relevant to how humans adapt to change and make choices in dynamic settings. For example, Rakow and Miler (2009) investigated repeated binary choice in which the associated probabilities of the outcomes could change over the sequence of trials. Specifically, the probability of one option would gradually change over a set of trials. The information given to participants was manipulated by providing the outcomes associated with each option or seeing a summary of the outcomes of previous trials. They observed a rapid adaptation (quick identification of the best option) when the probability changed, but a slow adaptation when only the outcome changed. The historic feedback helped but only in

early choices and not in later choices. Additionally, over all of their experiments, Rakow and Miler found some evidence that people react more quickly to negative changes than to positive changes. Their studies concluded with the importance of the adaptive nature of human memory and speculated how forgetting and recency of information can play an important role in adaptation.

Lejarraga, Dutt, and Gonzalez (2012) used Rakow and Miler's (2009) data to demonstrate how an Instance-Based Learning (IBL) cognitive model (Gonzalez et al., 2003) could account for that data. They compared the predictions from the IBL model to observed human choices suggesting that adaptation occurs through the reliance on recent outcomes. More recently, Lejarraga, Lejarraga, and Gonzalez (2014) investigated whether groups make better choices than individuals in dynamic tasks using similar problems with changing probabilities over time. They found that decisions made in groups were better than individual decisions in stable conditions, but groups were not superior to individuals after a sudden change had occurred in the probabilities. That is, groups had more difficulty in detecting and adapting to a sudden change compared to individuals. They also used an IBL model and a Bayesian updating model with "perfect memory" to explain why groups were slower at changing their policies compared to individuals.

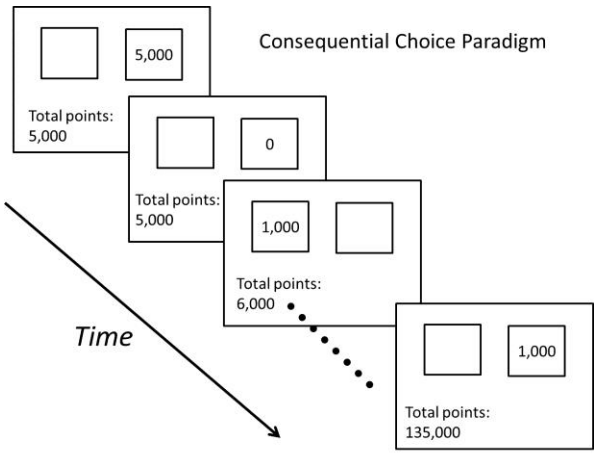

Figure 1: Repeated choice, consequential paradigm. The example shows that selecting the right button may result in an outcome of 1,000 with probability p=.2, 0 with p=.6 and 5,000 with p=.2, while the left button results in 1,000 with certainty (p=1). The probabilities are unknown to the participants.

We advance this line of research in four ways: First, we investigate how individuals adapt to *gradual* and *continual* change rather than to sudden changes. This is relevant to test the role of human memory in adjusting to gradual changes by the slightly altering past experiences. Second, we are also interested in looking at the *direction* of change. In research related to control of dynamic systems, researchers have found that adapting to change and being able to control a dynamic system in the long term, depends on whether the external environmental changes occur in a positive (i.e., increasing amounts) or negative (i.e., decreasing amounts) way (Gonzalez & Dutt, 2011). We present results from an experiment in a repeated choice paradigm in conditions of gradual positive and negative change. Third, we analyze exploratory behavior to evaluate how individuals explore the environment in order to detect gradual change. Fourth, we demonstrate the effects of human memory in these changing situations with an Instance-Based Learning model (Gonzalez et al., 2003), which relies on the ACT-R architecture's memory decay function (Anderson & Lebiere, 1998).

In addition to exploring gradual and continual change, the paradigm we implement involves high reward outcomes that change from very rare to near certain; analogous to a foraging animal in a changing environment once rich in resources that gradually deplete or vice versa (Mehlhorn et al., 2015). This design not only extends the results of Rakow and Miler (2009) to gradual and continual change but presents a test of boundary conditions in which exploration between options could be abandoned before change is detected due to very rare or very frequent early experiences. Accordingly, we expected better adaptation to change when the high risky outcome changes from very frequent to very rare.

Methods

Participants

Two-hundred and forty participants (88 Female, M_{age} =31.32) were randomly assigned to one of three conditions: increasing dynamic condition (N=76), decreasing dynamic condition (N=83), and stationary condition (N=81). Participants were recruited from Amazon Mechanical Turk for a "choice" experiment. They were paid $0.50 for participating and an additional bonus payment based on the points they accumulated over the course of the 100 trials, at a rate of 1 cent per 1000 points. The average bonus payment gained for the duration of the experiment of about 10 minutes was $0.26.

Design

The experiment asked participants to choose repeatedly between two options, with the goal of maximizing their long-term earnings that accumulate from each of the choices they make over 100 trials. The two options available in each choice include one risky option that could result in a high outcome (500 points) with probability p or a low outcome (0 points) with probability $1–p$, and a safe option that could result in a medium outcome (250 points) all the time. The presentation of the safe and risky options (left/right) was counter-balanced.

The main treatment involved the function of p which linearly increased, linearly decreased, or stayed stable as a function of time (choice trial number). In the increasing condition the probability of obtaining 500 points began at 0.01 and increased by 0.01 each trial, up to probability 1 at trial 100. This condition represents an environment where rewards change from extremely rare to certain as a function of time. In the decreasing condition the probability of obtaining 500 points started at 1 and decreased by 0.01 each trial, ending at 0.01 at trial 100. This condition represents an environment where rewards change from certain to extremely rare as a function of time. In the stationary

condition, the probability of obtaining 500 points stayed stable at 0.5 throughout.

Importantly, the cumulative expected value (EV) for all the options in all the conditions remains equal to 25,000 points over the course of the experiment. Thus, effectively selecting the risky option consistently over the course of 100 trials would result in approximately 25,000 accumulated points, the same overall accumulated outcome from selecting the safe option consistently over the course of 100 trials. However, the relative value of the two options changes over time. In the decreasing condition, the risky option is better than the safe option in the first 50 trials (i.e., it results in 500 points more often than 0 points) and then it becomes worse than the safe option in the last 50 trials. In the increasing condition, the risky option is worse than the safe option in the first 50 trials and it becomes better than the safe option in the last 50 trials. In the stationary condition the probability of getting 500 or 0 points in the risky option is always the same (0.5), so effectively the risky option is relatively as valuable as the safe option over the 100 trials.

Procedure

After providing consent and answering demographic questions, participants were given instructions for the game, and then they performed the choice task for 100 trials. Upon completion of the task, participants were given a final debriefing to determine whether they were aware of the changing probabilities and the direction of change.

Participants were given their total number of points accumulated, translated into a monetary bonus they earned, and then thanked for their participation.

Results

As a first step, we compared the proportion of risky choices (P-Risky) (see Figure 2, left panel). We analyzed the data using a generalized logit mixed-effects model with condition and block (blocks of 20 trials) as fixed effects, and subject-specific random intercepts. We found a significant difference in the proportion of risky choices participants made across conditions, $\chi^2(2) = 33.11$, $p < .001$. The P-Risky was higher in the decreasing condition (M = 0.48), followed by the stationary condition (M = 0.36), and the increasing condition (M = 0.25). There was also a significant effect of block, $\chi^2(4) = 229.09$, $p < .001$, and a significant interaction between condition and block, $\chi^2(8)$ = 1,289.77, $p < .001$.

Looking at Figure 2 (left panel), the trends over time reveal a decrease in the P-Risky for the stationary condition, suggesting a general tendency to gradually select the safe option over time, even when the options were objectively equal. This is explained by risk aversion (Kahneman & Tversky, 1979), which has been investigated in decisions from experience paradigms through IBL models (Lebiere, Gonzalez, & Martin, 2007). The frequency and recency of occurrence of the low outcome in the risky option creates an imbalance of preference towards the safe option (Lebiere et al., 2007).

Second, the patterns suggest that although the P-Risky in the increasing condition was the lowest, the overall proportion of risky choices moved in the direction of the increased probability in the non-stationary option. However, this adaptation seems to be slow. Initially, participants quickly favored the safe option as we observe from the immediate drop of the P-Risky in the first 10 trials, but they moved slowly towards preferring the risky option as per the increase in the probability of the high outcome.

Third, the P-risky in the decreasing condition reduced rapidly over the course of 100 trials. Initially, choices quickly favored the risky option but they started to favor the safe option more quickly as the probability of the high option decreases, suggesting probability matching behavior (Erev & Barron, 2005). We observe that people were faster to cross the P-Risky = 0.50 threshold in the decreasing condition compared to the increasing condition. In the increasing condition, P-Risky did not reach 0.50 until the 98th trial, whereas in the decreasing condition, P-Risky reached the 0.50 mark on the 50th trial, essentially tracking the probability function throughout (i.e., a demonstration of probability matching behavior; see also Rakow & Miller, 2009).

Participants seem to select the risky option that matches the probability of the high outcome. That is, participants seem to select the *maximizing* option more accurately in the decreasing rather than the increasing condition. To test this, we calculated the proportion of maximization choices (P-Max) before and after the objective change of the relative goodness of the options (e.g. trial 50; see Figure 3, left panel). We found a significant difference in the P-Max across increasing and decreasing conditions, $\chi^2(1) = 15.81$, $p < .001$. The P-Max was higher in the decreasing condition (M = 0.66) than the increasing condition (M = 0.57). In addition, participants maximized more in the first period of the task, $\chi^2(1) = 537.50$, $p < .001$, and the interaction was also significant, $\chi^2(1) = 1{,}916.19$, $p < .001$. The P-Max for the increasing condition was significantly higher in the first half (M = 0.83) than the second half (M = 0.33; $\chi^2(1)$ = 1,717.80, $p < .001$), whereas the order is reversed in the decreasing condition but to a lesser degree ($M_{first\ half}$ = 0.61, $M_{second\ half}$ = 0.64; $\chi^2(1) = 12.34$, $p < .001$). The contrast is quite stark: the maximizing rate never drops below 0.50 in the decreasing condition, but it is on average 0.30 in the second half in the increasing condition. This is consistent with the observation that people are adapting significantly more rapidly in the decreasing condition than in the increasing condition.

A possible explanation for the different degrees of adaptation between the increasing and decreasing conditions is the lack of exploration of the options. As we observe in the stationary condition, participants' choices drift towards the safe option over time, even when there is no change in probabilities and values. In the increasing condition, people might also have this tendency given that the safe option appears to provide higher payoffs more often than the risky option in the first few trials. This might prevent participants form exploring the risky option in later trials. In contrast, in the decreasing condition, since the risky option provides higher payoffs than the safe option in early trials, it is possible that people are more aware of the changes in the probability given that they are already selecting the risky option more often.

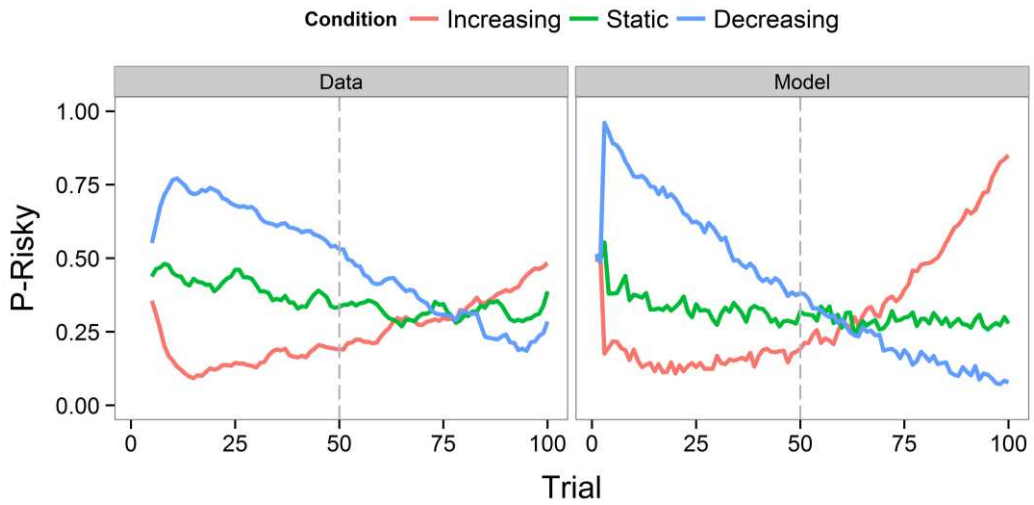

Figure 2: The left panel shows human responses and the right panel shows model predictions for each of the three conditions, marked in different colors, where the measure is the proportion of choosing the risky option across trials.

To test for exploration, we used a measure proposed in past research: the alternation rate (A-rate; Gonzalez & Dutt, 2011; 2012). This is the proportion of switches from one option to another in consecutive trials. We performed a similar analysis (mixed-effects logit model) and we found a significant effect of condition, $\chi^2(1) = 16.79$, $p < .001$, as participants switched more between options in the decreasing ($M = 0.30$) than the increasing condition ($M = 0.21$). The effect of period (before or after trial 50) was not significant, $\chi^2(1) = 0.43$, $p = .51$, but the interaction between condition and period was significant ($\chi^2(1) = 10.47$, $p = .001$): while there was a difference in A-rate between periods in the increasing condition ($p = .003$), this was not the case in the decreasing condition ($p = .12$; Figure 4, left panel).

Instance Based Learning Model

An IBL model designed to account for over-time effects of binary choice (Gonzalez & Dutt, 2011) is a generalist (it applies to a wide variety of tasks) instead of a specialist (a model that is made for one particular task; Lejarraga, et al., 2012) and it builds on the ACT–R cognitive architecture (Anderson & Lebiere, 1998). It proposes that a choice is a function of the accumulated value (*blended value*) for each of the two options, through experience. This value is a function of the outcomes observed and the associated probability of retrieving the corresponding instances from memory. Memory retrieval depends on the activation of a value that reflects how readily available this information is in memory. In this IBL model, activation reflects the *frequency* (how many times an outcome has been observed in the past), *recency*, and *noise* of the experience. The formulation of this model appears in multiple past publications (e.g., Gonzalez & Dutt, 2011; Lejarraga, et al., 2012), but for completeness we reproduce it here. A choice between the two options is made by using the blended value V which represents the value of option j in a particular trial t:

$$V_{jt} = \sum_{i=1}^{n} P_{it} x_i$$

where x_i refers to the payoff obtained in each option stored in memory as instance i for the option j, and p_i is the probability of retrieving that instance from memory, which is relative to the activations of other instances in option j:

$$P_{jit} = \frac{e^{A_{it}/\tau}}{\sum_j e^{A_{jt}/\tau}}$$

where τ is random noise defined as $\tau = \sigma\sqrt{2}$, and σ is a free noise parameter. The activation of instance i represents how readily available the information is in memory:

$$A_{jit} = \sigma \ln\left(\frac{1-\gamma_{jit}}{\gamma_{jit}}\right) + \ln \sum_{t_p \in \{1,\ldots,t-1\}} (t - t_p)^{-d}$$

The activation is higher when instances are observed frequently and more recently. When an instance is not observed often, the memory will decay with the passage of time (the parameter d, the decay, is a non-negative free parameter that defines the rate of forgetting). The noise component σ is a free parameter that reflects noisy memory retrieval, γ is a random sample from a uniform distribution (between 0 and 1), and t_p denotes all the previous trials that outcome i was observed.

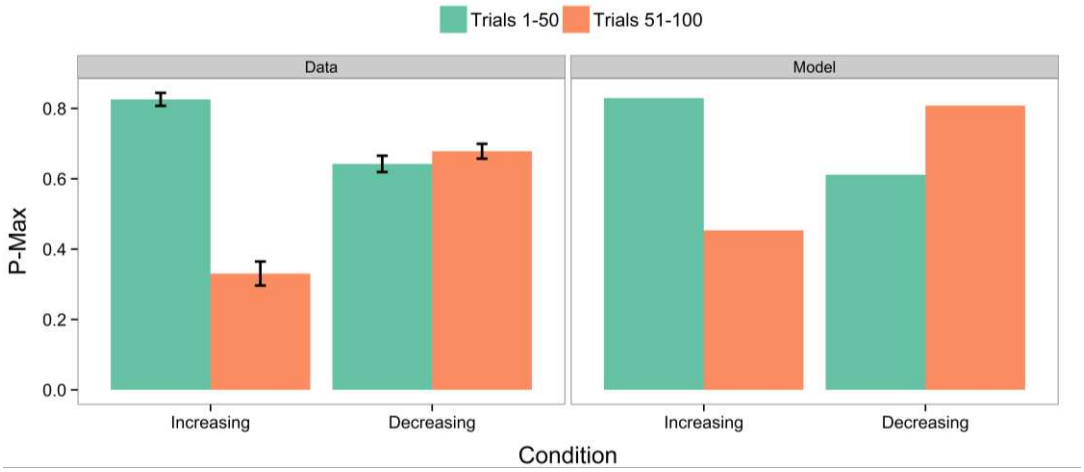

Figure 3: The left panel shows human data and the right panel shows model predictions of average maximization rates (P-Max) in the changing probability conditions (increasing and decreasing) before and after the halfway point (50th trial).

We produced predictions via simulations using this IBL model with parameters from past research (d = 5 and σ = 1.5; see Lejarraga, et al., 2012). We simulated the choices over time of 500 participants in each of the experimental conditions. The right panels in Figures 2, 3, and 4, show the model predictions that correspond to each of the results observed from the human data.

The model made predictions of choice behavior across time that reflected similar trends in human data (right panel, Figure 2). The model differed in how often it chose the risky option on average across the three conditions. In the decreasing condition, the mean P-Risky over the 100 trials was 0.40; in the increasing condition the mean P-Risky was 0.32; in the stationary condition it was 0.33. This was largely due to worse adaptation in the increasing condition, in which P-Risky does not reach 0.50 until the 77th trial (so 27 trials after it would have been beneficial to do so). On the other hand, in the decreasing condition, the model begins choosing the safer alternative in advance and P-Risky crosses the 0.50 mark on the 34th trial. In the stationary condition, P-Risky drops to around 0.30 and remains around that level. In agreement with the observation in human participants, although the two options have the same EV, the model chooses the safe option about two times more on average than the risky option.

The P-Max between conditions in the first and second half of the experiment is shown in Figure 3 (right panel). The average P-max in the first half (M = 0.83) is higher than in the second half (M = 0.48). However, in the decreasing condition we observed a trend in the opposite direction. In the decreasing condition we find that the average P-Max in the first half (M = 0.60) is lower than the second half (M = 0.81). Regarding A-rate (Figure 4), the model accurately predicts more switching in the decreasing condition (M = 0.28) compared to the increasing condition (M = 0.23).

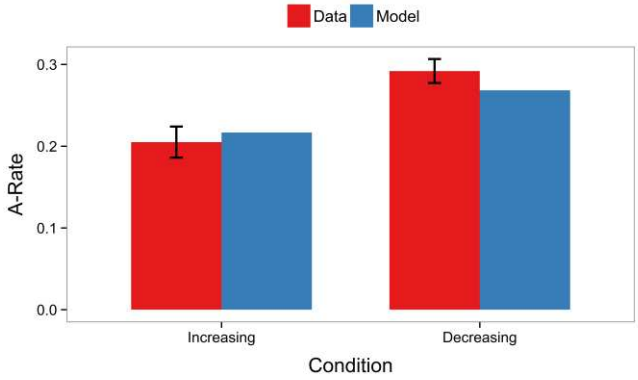

Figure 4: Average observed (Data) and predicted (Model) alternation rates (A-rate) in the changing probability conditions (increasing and decreasing).

In contrast to human participants, the predictions of the model are more extreme: adapting better in the decreasing than in the increasing condition and doing better in terms of maximizing choices in the second half than in the first. However, these are good predictions given that this is an "out of the box" model prediction, where the simulated data were produced in the complete ignorance of human data. In fact, when we calculated the mean squared difference (MSD) for each condition, we found that the predictions in the stationary condition were the closest to observed data, with an MSD of 0.006; the decreasing condition was the next closest with an MSD of 0.018, and the increasing condition was the next one with an MSD of 0.020.

Discussion

The main purpose of the current investigation was to examine how people adapt their choices to gradual and continual change of event probabilities. Specifically, we were interested in whether the direction of change

(increasing or decreasing probabilities of maximum outcomes) would have an effect on people's choice behavior. According to economic theory one should expect no difference in choice across conditions, since the cumulative EV of all options was *the same* in all experimental conditions. In contrast, as expected by cognitive theories, results show that people are sensitive to the dynamics of experienced outcomes and to the direction of change of the associated probabilities to these outcomes.

Three main phenomena emerged from this investigation: 1) risk aversion in experience, 2) slow adaptation to increasing probabilities, and 3) fast adaptation to decreasing probabilities. These patterns of risky choice are reinforced by two observed behaviors: the maximization and the alternation behavior. In the increasing condition, people chose the best option in the first half of their experience but they fell far below the average optimal behavior in the second half; while in the decreasing condition participants stayed above average optimal behavior throughout. The results suggest that participants explore the two options more in the decreasing than the increasing condition.

We observe that the IBL model provides close predictions to the observed behavior. Even though the performed simulations were not exhaustively in line with what was observed in the task, the model provided insightful observations into the mechanisms of adaptation to change. Introspecting into the IBL model's mechanisms we observe, and it is also discussed in more detail in Lebiere et al. (2007), that the model naturally develops a preference for the safe option rather than the risky option. This is due to the experiences of extreme outcomes in the risky option and the blending choice mechanism of the model that "blends together" these outcomes, giving rise to a slight preference for the safe option (i.e., the stationary option). These predictions emerge from the activation of instances that reflects the frequency and recency of the occurrence of outcomes. The stationary outcome develops initially a higher probability of retrieval and a slightly higher blended value. In the increasing condition, this tendency prevents the model from exploring the risky option, resulting in "lack of awareness" of the change. In contrast, in the decreasing condition, because the low outcome of the risky option has an extreme low probability of occurrence early on, the model develops a preference for the risky over the safe option. This results in "awareness" of the change in the probability which helps the model being more successful at adapting to the probability decrease.

In conclusion, the area of dynamic experience-based decision-making has remained largely unexplored and this study attempted to provide a deeper understanding of the factors that are involved in the adaptation to continuous dynamic change. We found that people were slower at adapting to changes in the outcome probability when a high outcome changes from rare to frequent compared to a high outcome changing from frequent to rare. People are slow at switching to a risky choice in the increasing condition and fast at switching to a safe option in the decreasing condition. Differences in exploration of the available options, joined with the dynamics of experience and the cognitive effects involved (frequency and recency of experiences) provide an explanation of this behavior.

Acknowledgements

This work was supported by the National Science Foundation Award Number: 1154012 to Cleotilde Gonzalez.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16*, 215-233.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review, 69*, 109-135.

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

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591-635.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. P*sychological Review, 118*, 523-551.

Gonzalez, C., & Dutt, V. (2012). Refuting data aggregation arguments and how the IBL model stands criticism: A reply to Hills and Hertwig (2012). *Psychological Review*, *119*, 893-898.

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.

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

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

Lebiere, C., Gonzalez, C., & Martin, M. (2007). *Instance-based decision making model of repeated binary choice.* Paper presented at the 8th International Conference on Cognitive Modeling, Oxford, UK.

Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making, 25*, 143-153.

Lejarraga, T., Lejarraga, J., & Gonzalez, C. (2014). Decisions from experience: How groups and individuals adapt to change. *Memory & Cognition, 42*, 1384-1397.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*, 191–215.

Rakow, T., & Miler, K. (2009). Doomed to repeat the successes of the past: History is best forgotten for repeated choices with nonstationary payoffs. *Memory & Cognition, 37*, 985-1000.