The Influence of Risk Aversion on Visual Decision Making

Ruixin Yang (r4yang@cs.ucsd.edu)
Department of Computer Science and Engineering
9500 Gilman Drive
La Jolla, CA 92093 USA

Garrison W. Cottrell (gary@ucsd.edu)
Department of Computer Science and Engineering
9500 Gilman Drive
La Jolla, CA 92093 USA

Abstract

The ability to decide between multiple fixation targets in complex visual environments is essential for our survival. Evolution has refined this process to be both rapid and cheap, allowing us to perform over 100,000 saccades a day. Previous models for visual decision making have focused on maximizing reward magnitude or expected value (EV = probability of reward × magnitude of reward). However, such methods fail to incorporate utility, or happiness derived from reward, optimizing strictly on nominal reward values. We propose an alternative model for visual decision making, maximizing utility as opposed to value under the assumption of a decreasing marginal utility curve. To test our model, we asked 10 UCSD graduate students to participate in an eyetracking experiment where they choose between different fixation targets presented on a brief display. The reward for each target was generated from fixed, predetermined distributions with different variance that was initially unknown to the subjects. The subjects were asked to maximize their reward for each test session within the experiment. Comparing our results with expected value and reward optimizing hedge algorithms, we show that utility-based models more accurately reflect human behavior in visual decision making tasks.

Keywords: Visual decision making; risk aversion; utility theory; reward.

Introduction

Target selection is a complex optimization task that the human visual system must complete thousands of times per day. Assuming a probabilistic, stationary distribution for reward, the problem is directly reducible to the multi-armed bandit problem (Lai & Robbins, 1985; Freund & Schapire, 1997; Auer et al., 2003; Chaudhuri, Freund & Hsu, 2009). Despite the complexity of the problem, an efficient, low-cost algorithm is necessary to permit rapid saccades to be made in the noisy, low-resolution perceptual environment. Previous attempts to model visual decision making involve a probabilistic, reward magnitude framework where the probability of fixation is weighted based on the expected value (EV), defined as the probability of reward multiplied by the magnitude of reward (Milstein & Dorris, 2007; Navalpakkam et al., 2010; Platt & Glimcher, 1999). Depending on the approach, the definition of the probability of reward can be taken either from the traditional economic context as the probability of obtaining a reward upon target fixation (Milstein & Dorris, 2007; Milstein & Dorris, 2011; Platt & Glimcher, 1999) or from the perspective of noisy sensors, modifying the probability of a target's location given the noise (Navalpakkam et al., 2010). Alternative approaches that have achieved comparable accuracy to expected value strategies include a study on rhesus monkeys by Milstein and Dorris (2011), where they show reward magnitude alone may explain saccadic preparation and reaction time data.

Despite the success of expected value models, there are many real life examples where behavior does not maximize expected value. For example, for many types of large investment (e.g. automobile, home, healthcare), there is a set of insurance policies to reduce risk. Insurance companies sustain themselves by making a small profit while reducing risk for consuming individuals. If it were the case that every individual viewed reward maximization as maximizing their expected value, these institutions would no longer be profitable and would cease to exist. However, there are many instances where individuals are willing to disproportionally sacrifice some of their assets to reduce the probability of an extremely undesirable outcome. As a result, insurance is often viewed as mutually beneficial and is encouraged in many situations. Bernoulli (1738/1954) provided the first examples of deviation from expected value behavior, stating that decisions should be based on an individual's current wealth, with less wealthy individuals being more averse to risky decisions. Brocas and Carrillo (2009) presented a simple illustration where two perfectly rational individuals may come to opposite conclusions in situations of uncertainty due to differences in utility preferences. The goal of our experiment is explore the concept of utility maximization and risk aversion in the context of visual decision making, where decisions are made in quick succession, without much time for the conscious evaluation of value.

We provide an alternative model for visual decision making that attempts to maximize utility, or happiness derived from reward, rather than expected value. In the following sections, we will formally define our model, as well as provide a mathematical justification for why expected value fails to account for behavior with respect to most types of reward. We also compare the predictions of our model with those made by the expected value model and

two well-known machine learning algorithms. Our results show that a risk averse, utility maximization model performs significantly better than the other models at describing human eye movement behavior under all experimental conditions.

Model Description

Assume that there are n targets: x_i, where $i = 1, \ldots, n$.

- Let $\pi_{i,l}$ be the probability of encountering target x_i at location l.
- Let v_i be the value for fixating on target x_i.

The expected value for fixating at location l can be calculated as follows:

$$EV(l) = \sum_i \pi_{i,l}\, v_i. \qquad [1]$$

However, value alone is often a poor measurement for reward (Bernoulli, 1954; von Neumann & Morgenstern, 1953). This is due to the fact that people tend to have a decreasing marginal utility for most types of reward.

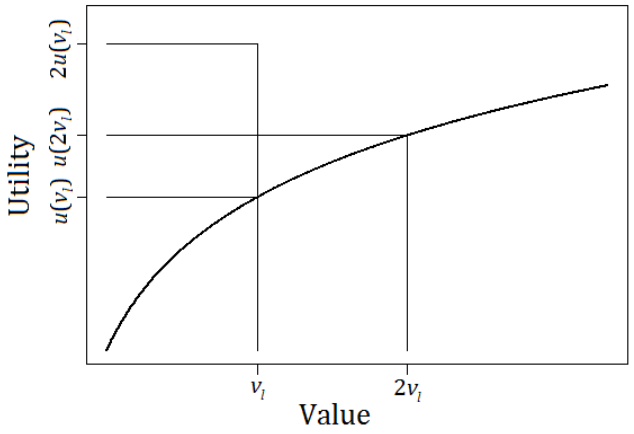

Figure 1: Individuals tend to have a decreasing marginal utility for most types of reward due to priorities in reward allocation (see text). Note that for decreasing marginal utility, the utility derived from doubling the value of reward is less than twice the utility of the original value: $u(2v_l) < 2u(v_l)$.

Intuitively, decreasing marginal utility arises from how consumption of reward is allocated. The first units of a reward such as money tend to be spent towards essential necessities, while later units tend to be used on luxury goods. Similar arguments could be made for other rewards such as food, shelter, and material goods. Figure 1 depicts a standard decreasing marginal utility curve.

One implication of holding a decreasing marginal utility curve is that the risk averse decision will frequently maximize utility. For example, Figure 2 shows that under the decreasing marginal utility assumption, an individual will always prefer an action that generates a guaranteed

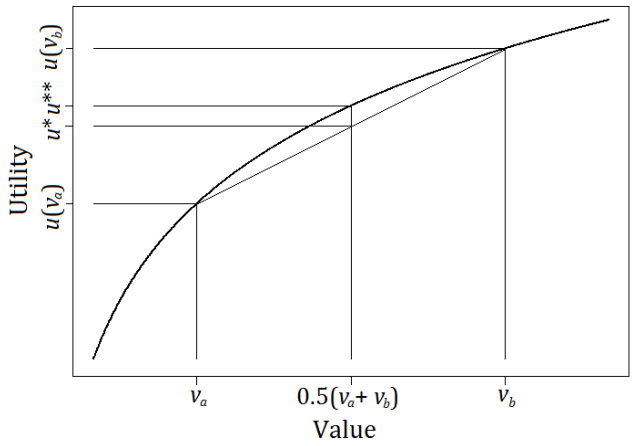

Figure 2: A property of the decreasing marginal utility curve is that individuals should demonstrate risk averse behavior in their decision making. Despite having identical expected values, a fair chance between v_a and v_b yields u^*, which is less than the utility provided by $0.5(v_a + v_b)$, which yields u^{**}. Note that the choice between v_a and v_b has greater variance than the single point, $0.5(v_a + v_b)$.

reward to one which provides a fair gamble between two outcomes that preserves expected value. This result generalizes: individuals with a decreasing marginal utility for reward should always prefer the choice with lower variance, given options of equal expected value.

Replacing value (v_i) with utility $u(v_i)$, we obtain the following equation for expected utility:

$$EU(l) = \sum_i \pi_{i,l}\, u(v_i). \qquad [2]$$

The goal of the subject is to find the location in the scene that contains the target which maximizes expected utility. Therefore, the objective function for maximizing utility is:

$$f = \text{argmax}_l \sum_i \pi_{i,l}\, u(v_i). \qquad [3]$$

To model $u(v_i)$, we note that it is monotonically increasing, but has a decreasing slope leading to diminishing returns with increased reward. For our model, we represent the curve using a natural logarithm because of its simplicity and because it shares the same properties as the utility curve. It is important to note that no two individuals hold the same utility function for reward, and the natural logarithm reflects a hypothetical utility function of the average individual.

The new objective function thus becomes:

$$f = \text{argmax}_l \sum_i \pi_{i,l} \ln(cv_i + 1) \qquad [4]$$

where c is a constant value reflecting the magnitude of risk aversion. In our experiment, v_i is a value that must be learned by the subject for each experimental phase.

Based on feedback, in each trial we approximate $u(v_i)$ as a weighted average across past observed rewards, $\ln(cr_1 + 1), \ln(cr_2 + 1), \ldots, \ln(cr_t + 1)$. Similarly, v_i is approximated by the weighted average of the observed rewards sequence $(r_1, r_2, \ldots, r_t)$. Since the targets in our experiment are visually dissimilar, $\pi_{i,l}$ becomes close to 0 or 1 depending on the target. This allows us to approximate $\pi_{i,l} \in \{0, 1\}$ and simplifies our calculation.

Experimental Methods

Subjects

Ten naive subjects participated in the experiment after providing informed consent. All subjects were right-handed graduate students from the University of California, San Diego Computer Science and Engineering department.

Experimental Procedure

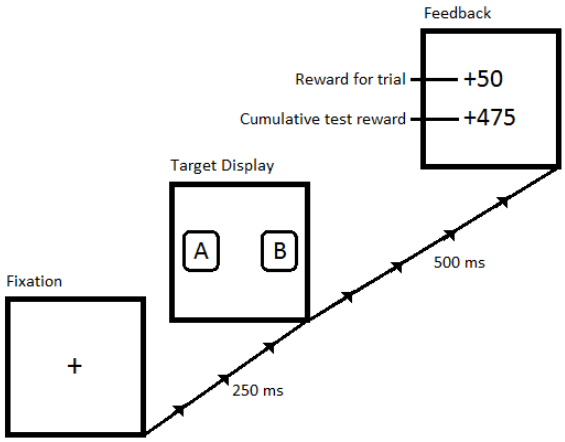

Figure 3: The basic experimental setup. At start of every trial, subjects were asked to hold a central fixation for 250 ms before being presented the target display. Once the targets are presented, the subject must make a saccade to one of the targets within 500 ms, when feedback is displayed. If no decision is made within the 500 ms, the subject receives no reward for the trial and is notified that they must make their decision faster. Subjects are permitted to spend as long as they wish on the feedback screen prior to starting the next trial to make adjustments to their strategy.

We model our experiment using similar parameters to those used in Navalpakkam et al. (2010). At the beginning of every trial, subjects were asked to indicate they were ready by fixating at a center fixation point and pressing the "enter" key on the keyboard. Each trial begins with a central fixation '+' presented for 250 milliseconds. Subjects indicate their choice by saccading to one of the targets. Then, subjects were presented with an image containing two targets labeled 'A' and 'B' for 500 milliseconds. Figure 3 provides an illustrated description of each trial.

Targets in the experiment appeared at 7 degrees eccentricity from the central fixation point, and were horizontally aligned. The target stimuli were 1.8 degree in height and was each encompassed by a 3.6 × 3.6 degrees square border. The location where each target appeared was randomly generated from trial to trial. Subjects viewed the display on a 19-inch cathode ray tube (CRT) monitor at a distance of 30 inches from the screen.

The experiment consists of an initial training phase and three test phases of 50 trials each. In the training phase, the rewards for the two targets were drawn from discrete uniform distributions, U[0, 50] and U[50, 100]. Subjects were required to learn which target yielded the higher reward and fixate on that target for at least 75 percent of the training trials before being allowed to proceed to the test phase. This was done to ensure the subject was accustomed to making quick and accurate saccades while wearing the eyetracking device. Subsequently, each test phase consisted of two targets of equal mean, but different variance.

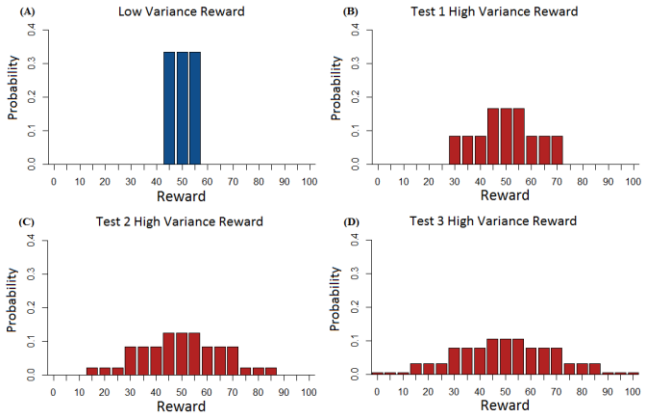

Figure 4: The distribution of target reward for each test. Of the two targets, the lower variance target (4A) was kept constant for all three tests, while the distribution of the higher variance targets (4B), (4C), and (4D) changed for Tests 1, 2, and 3 respectively.

The reward for each target was generated from the distributions described in Figure 4. For each test, the higher variance target's distribution was adjusted, while the lower variance distribution was kept constant across all of the tests. Subjects did not know the identity or probability distribution of the target at the beginning of each trial or test phase in the experiment. They were instructed to learn the distributions and maximize their reward for each test phase.

We used an Eyelink 1000 eyetracker from SR Research to record the subjects' eye movements. At the beginning of each test, the eyetracker was recalibrated using a nine-point calibration across the edge and center of the display.

Modeling Procedure

Performance for each model was measured by the percentage of trials in which the model matches the human

fixation decision. For each trial, the utility-based risk averse model picked the location that maximized the objective function described in Equation [4], while the expected value model maximized $\text{argmax}_l \sum_i \pi_{i,l} v_i$. We chose the constant c from Equation [4] by binary search across the [0, 10] interval, finding the value of c that best predicted the human subject data. However, the results were not sensitive to the exact choice of c.

Every model prediction was compared with the decision made by the subject. Updates to both models (i.e. running averages to v_i and $u(v_i)$), were done based on the experimental feedback provided to the subjects for each individual trial, as if the model had made the same choice as the subject. In addition to the greedy algorithms, where the model decision is always the one that maximized the objective function, we compared the performance of three policies that include exploration of less-valued alternatives: ε-greedy, decreasing ε, and softmax (Sutton & Barto, 1998). We tested epsilon and softmax temperature values of 0.05, 0.10, and 0.20. The exploration algorithms may be summarized subsequently as follows:

ε-greedy:

Initialize ε
For every trial
  Generate random number r ∈ [0, 1]
  If r > ε
    Take action maximizing the objective function
  Else
    Randomly generate action from uniform distribution

Decreasing ε:

Initialize ε, dec_rate = ε / (num_trials – 1)
For every trial
  Generate random number r ∈ [0, 1]
  If r > ε
    Take action maximizing the objective function
  Else
    Randomly generate action from uniform distribution
  Update ε = ε – dec_rate

Softmax:

Initialize τ
For every trial
  Generate action *i* based on probability density:

$$\frac{e^{Q_t(i)/\tau}}{\sum_{j=1}^{2} e^{Q_t(j)/\tau}}$$

  where $Q_t(i)$ is the running average of reward for choosing *i*.

Aside from comparing prediction performance against expected value, we also compared our results against two well-known machine learning algorithms, Hedge (Freund & Schapire, 1997) and Normal-Hedge, (Chaudhuri, Freund & Hsu, 2009) from the multi-arm bandit problem literature. Both algorithms are designed to maximize reward, given a single parameter value (β for Hedge and c_t for Normal-Hedge). Before we continue, it is important to take note of one subtle difference in the objective function of the multi-armed bandit problem with our problem. The objective function of the multi-armed bandit minimizes regret (defined as the difference in reward between the ideal and the chosen action) as opposed to maximizing accumulated reward. To address this, we linearly transformed the reward obtained to range from [0, 1] and compute the loss of each action as the difference, 1 - reward. In our Hedge implementation, we tested the entire range of temperature values [0, 1] in increments of 0.05. We find the best temperature setting to be $\beta = 0.05$, and use it for our analysis. For Normal-Hedge, the algorithm is self-adapting around a variable constraint c_t (note this variable is unrelated to the variable c from Equation [4]). We solve for c_t using line search as recommended by the authors. A detailed description of the algorithm as well as the proof on performance bounds may be found for Hedge in Freund & Schapire (1997) and Normal-Hedge in Chaudhuri, Freund & Hsu (2009).

In all our models, we excluded two subjects from our experiment as their data lay beyond two standard deviations from the mean number of saccades to the lower variance target. Note that we did not purposely remove risk seekers as this is equivalent to removing data with respect to the higher variance target due to the fact that there are two targets. Of these, one of the subjects was removed because he systematically fixated only at the target that appeared on the left side of the display, regardless of the identity of the target.

To address potential location bias concerns in making saccades, we recruited only right-handed subjects for our study. In addition, we tested our models with and without two location-based prior probabilities obtained from subject responses. The priors were the probability of fixating at each potential target location, and the probability of returning, given a previous saccade to the same location on the previous trial. In all of our models, there was no significant change in performance when we incorporated the priors under a Bayesian setting. For this and all model comparisons used in this paper, we used a paired t-test to compare the performance between models.

Results

Behavioral Data

The results of the human experiment are presented in Figure 5. For each test phase, we maintained a record of the number of saccade decisions to each target. The reward distribution for higher variance target in Tests 1-3 shared the same mean, but differed in variance as shown in Figure 4 (B-D) respectively, so the utility for these choices will increase. The lower variance target maintained the same distribution for all three tests. In Test 3, the subjects showed significantly risk averse behavior (p = 0.0321), choosing the lower variance target 54.8 percent of the time. As the difference in variance between the targets decreased, subjects

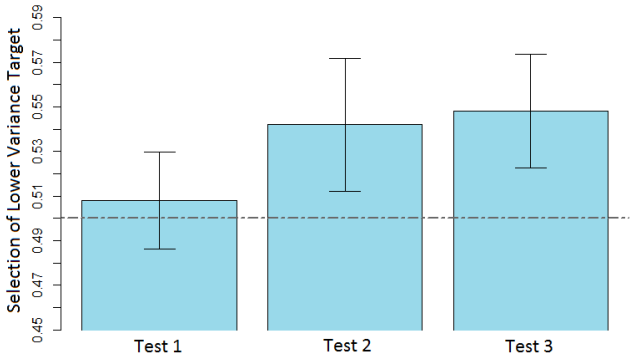

Figure 5: The results from the human experiment. The gray line represents indifference between the two targets. The reward distribution for the higher variance target in Tests 1-3 shared the same mean, but differed in variance as shown in Figure 4 (B-D) respectively. Of the three tests, subjects showed significant risk averse behavior in Test 3 (the test with greatest difference in variance between the two targets), where they chose the lower variance target 54.8% of the time.

became increasingly indifferent between the two targets. Subjects in Test 2 chose the lower variance target 54.2 percent of the time ($p = 0.0861$), while subjects in Test 1 chose the lower variance target 50.8 percent of the time ($p = 0.621$).

Choosing a Value for c

Recall from Equation [4] in the model description where a constant, c, was included to allow for fine-tuning of the magnitude of risk-averse preferences. One interesting, and perhaps surprising result is that most reasonable settings of c outperform the expected value model in predicting human behavior. For this reason, we simply chose a local maximum using a binary search across positive values of c. Exceptions to this include $c = 0$ (when the strategy reduces to random) and extremely large values of c (when most rewards share approximately the same value).

Comparison with Expected Value

We simulated the expected value and utility-based risk averse strategies for 100 simulations (c = 2.48) using greedy, ε-greedy, decreasing ε, and softmax exploration functions (Sutton & Barto, 1998). Our results show that all non-greedy algorithms perform significantly worse than their greedy counterparts ($p < 0.001$). Simple exploration strategies yield poor performance because although they are capable of accurately capturing the probability of exploration, they fail at correctly predicting the trials on which they occur. As a result, since the probability of performing a reward maximizing action for any given trial is greater than the probability of

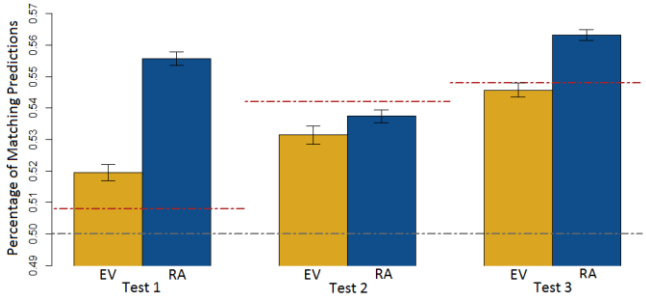

Figure 6: A comparison of the model fits between the greedy expected value and greedy utility-based risk averse (RA) strategies in predicting human data. In all three test conditions, the utility-based risk averse strategy significantly outperformed the expected value (EV) strategy for T = 100 simulations ($p < 0.001$). The gray dotted line represents chance performance, while the red dotted line represents fit obtained by always choosing the lower variance target (an omniscient model). Given a limited history of reward, the subject may choose the higher variance target as a result of a greedy action. Performance above the red dotted line suggests that the algorithm was fairly accurate its prediction of when the subject chose to take such greedy action as opposed to exploring the other option.

exploration, the greedy version of the algorithm will significantly outperform their exploration counterpart.

Figure 6 compares the utility-based risk averse strategy with the expected value strategy in predicting human behavior. The results show that although both models perform significantly above chance, maximizing across utility significantly outperforms value maximization ($p < 0.001$) for all three test conditions. The red dotted line provides a benchmark for how much performance may be obtained from a strategy defined by choosing only the lower variance target. Note that this is an overprediction, since at the beginning of each test phase, the subject does not know which of the two targets holds lower variance (or even the value of their reward).

Comparison with Hedge

We compare the fit of the utility-based risk averse strategy with two well-known algorithms for solving the multi-armed bandit problem from machine learning. We choose to implement hedge algorithms over an alternate strategy, Exp4 (Auer et al., 2003) due to its superior performance under conditions where the reward distribution is fixed (the reward probability distributions are fixed for our experiment as shown in Figure 4). We test twenty values for the temperature value (β) in Hedge and present the result for the optimal setting, $\beta = 0.05$. For Normal-Hedge, we find the constraint, c_t, via line search as recommended by the authors. Figure 7 shows a summary of our results. In all three test conditions, the

utility-based risk averse strategy significantly outperforms hedge algorithms under their optimal settings ($p < 0.001$).

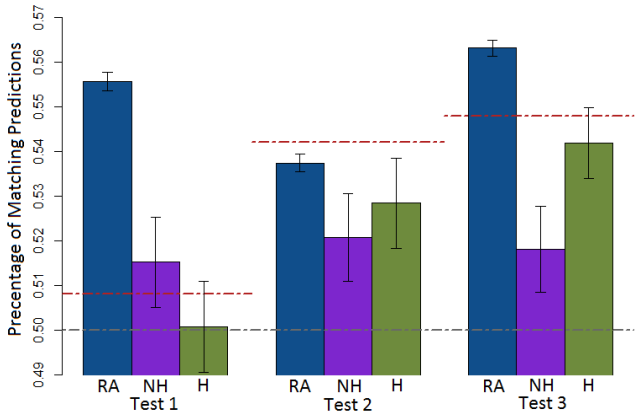

Figure 7: The results comparing the performance between the utility-based risk averse strategy with Hedge (H) and Normal-Hedge (NH) algorithms on predicting human data. In all three test conditions, the utility-based risk averse strategy significantly outperformed the Hedge and Normal-Hedge strategies for T = 100 simulations ($p < 0.001$). The gray dotted line represents chance performance, while the red dotted line represents performance obtained from only choosing the lower variance target.

Discussion and Future Work

Our work shows that by constructing models from a utility maximization standpoint, we are able to make predictions regarding human behavior that would otherwise be impossible in situations involving risk. Previous models in saccadic prediction involved a direct integration of the probability and magnitude of reward, ignoring risk derived from variance in reward distributions. In this paper, we present evidence suggesting the importance of such parameters when modeling visual decision making. Our findings show that under conditions of uncertainty, the human visual system takes a risk averse approach, taking account of the variance of the reward distribution in addition to the mean.

However, the current utility-based risk averse model does not address all questions that arise with the incorporation of risk. In particular, the work does not address issues raised from prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Kusev et al., 2009). For example, in the context of the experiment, there is no loss associated with viewing any target, and thus the asymmetry between loss and gain perception could not be modeled. Likewise, there are many situations where risk seeking behavior is exhibited and is the utility optimizing choice. While both conditions may arise in vision, prospect theory could not be modeled under the current experimental framework, and risk seeking behavior would require a change in the shape of the utility function. However, despite these limitations, we believe that the current work presents a starting point for analyzing visual decision making under uncertainty.

Acknowledgements

We would like to thank the members of the GURU research lab, and the Perceptual Expertise Network for comments and feedback on this work. The work was supported in part by NSF Grant #SBE0542013.

References

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. (2003). The non-stochastic multi-armed bandit problem. *SIAM Journal on Computing*, *32*, 48-77.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23-36.

Brocas, I., & Carrillo, J. (2009). Information acquisition and choice under uncertainty. *Journal of Economics and Management Strategy*, *18*, 423-455.

Chaudhuri, K., Freund, Y., & Hsu, D. (2009). A parameter-free hedging algorithm. *Advances in Neural Information Processing Systems*, *22*, 297-305.

Freund, Y., & Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*, 119-139.

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

Kusev, P., van Schaik, P., Ayton, P., Dent, J., & Chater, N. (2009). Exaggerated risk: prospect theory and probability weighting in risky choice. *JEP:LMC*, **35**: 1487-1505.

Lai, T. L., & Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, *6*, 4-22.

Milstein, D., & Dorris, M. (2007). The influence of expected value on saccadic preparation. *Journal of Neuroscience*, *27*, 4810-4818.

Milstein, D., & Dorris, M. (2011). The relationship between saccadic choice and reaction times with manipulations of target value. *Frontiers in Neuroscience*, *5*.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *PNAS*, **107**: 5232-5237.

Platt, M., & Glimcher, P. (1999). Neural correlations of decision variables in parietal cortex. *Nature*, **400**: 233-238.

Sutton, R., & Barto, A. (1998). *Reinforcement learning*: *An introduction*. MIT Press, Cambridge, MA.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of risk and uncertainty*, *5*, 297-323.

von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton University Press, Princeton, NJ.