Modeling Subjective Experience-Based Learning under Uncertainty and Frames

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

Citation

As Published
http://dl.acm.org/citation.cfm?id=2893873.2893925

Publisher
AAAI

Version
Author's final manuscript

Accessed
Thu Dec 06 03:05:17 EST 2018

Citable Link
http://hdl.handle.net/1721.1/109317

Terms of Use
Creative Commons Attribution-Noncommercial-Share Alike

Detailed Terms
http://creativecommons.org/licenses/by-nc-sa/4.0/
Modeling Subjective Experience-based Learning under Uncertainty and Frames

Hyung-il Ahn  
IBM Research  
650 Harry Road, San Jose, CA 95120  
hiahn@us.ibm.com

Rosalind W. Picard  
MIT Media Lab  
75 Amherst St, Cambridge, MA 02139  
picard@media.mit.edu

Abstract

In this paper we computationally examine how subjective experience may help or harm the decision maker’s learning under uncertain outcomes, frames and their interactions. To model subjective experience, we propose the “experienced-utility function” based on a prospect theory (PT)-based parameterized subjective value function. Our analysis and simulations of two-armed bandit tasks present that the task domain (underlying outcome distributions) and framing (reference point selection) influence experienced utilities and in turn, the “subjective discriminability” of choices under uncertainty. Experiments demonstrate that subjective discriminability improves on objective discriminability by the use of the experienced-utility function with appropriate framing for a given task domain, and that bigger subjective discriminability leads to more optimal decisions in learning under uncertainty.

Introduction

There are two seemingly contradictory experimental results regarding the role of subjective experience in human learning and decisions under uncertainty: Iowa gambling experiment (Bechara and Damasio 2005; Bechara et al. 1997; Naqvi, Shiv, and Bechara 2006). For Iowa gambling tasks, it should be noted that the optimal option involved safer gain outcomes, whereas the suboptimal option involved risky outcomes with long-run expected loss.

Second, Shiv et al.’s experiment (Shiv et al. 2005) in which choices are made between the option with higher mean and more uncertain outcomes (option 1) vs. the option with lower mean and less uncertain outcomes (option 2) (e.g., Domain 2 in Figure 1) has presented the harmful side of subjective emotional learning in terms of optimal decision behavior. In Shiv et al.’s actual experiment, the task involved 20 rounds of investment decisions between the optimal option with risky outcomes (investment, $3.5 gain with 50% chance and $1 loss with 50% chance for a choice, expected return = $1.25) and the safer suboptimal option (no investment, $1 gain for sure each choice, expected return = $1). Here also, normal people tended to select the option involving safer gain outcomes (but suboptimal in this task) more often than patients with emotional deficits.

In this paper we computationally explain how and when subjective experience (subjective discriminability) can lead to more or less optimal learning than objective experience in view of the interaction of framing and task domain. Our work contributes a novel unified framework that explains both the Iowa experiment, Shiv et al.’s experiment, and a variety of decision making tasks in this perspective.

In our view, both Iowa and Shiv et al.’s experiments illustrate that normal people tend to have uncertainty-averse and loss-averse attitude when they are faced with potential consistent gains. Furthermore, the task domain (underlying outcome distributions), interacting with the given gain frame is one factor that determines whether people’s subjective experience and uncertainty aversion help or harm their optimal decision making and learning under uncertainty.

We propose that, provided the decision maker’s representative risk attitude in each frame (gain or loss frame), the role of subjective experience-based learning depends on the task domain, the frame (reference point) selected by the decision maker’s emotional deficits related with the ventromedial prefrontal cortex (vmPFC) damage are not (Bechara and Damasio 2005; Bechara et al. 1997; Naqvi, Shiv, and Bechara 2006). For Iowa gambling tasks, it should be noted that the optimal option involved safer gain outcomes, whereas the suboptimal option involved risky outcomes with long-run expected loss.

The outcome distributions actually involved in IOWA and Shiv’s experiments were not Gaussian. However, Domain 1 and Domain 2 in Figure 1 represent the essential characteristics of those distributions in a mathematically simple way.

1The outcome distributions actually involved in IOWA and Shiv’s experiments were not Gaussian. However, Domain 1 and Domain 2 in Figure 1 represent the essential characteristics of those distributions in a mathematically simple way.

2If the decision maker’s own reference point for evaluating outcomes is smaller (or greater) than most sampled outcomes; thus,
decision maker themselves, and their interaction.

We investigate how domains and frames influence subjective experience and in turn, the “subjective discriminability” of choices. The concept of discriminability (Thurstone 1927; Busemeyer and Townsend 1993) characterizes the level of easiness in figuring out which option is optimal with fewer trials; thus, the discriminability is a key factor in regulating the trade-offs between exploration and exploitation and quickly detecting the optimal decision in learning. To model subjective experience, we propose the “experienced-utility function” based on a prospect theory (PT)-based parameterized subjective value function (Figure 2) (Kahneman and Tversky 1984; Kahneman 2003).

Using two-armed bandit task simulations, we compare subjective discriminability from the experienced-utility function (utility = PT-based subjective value) with objective discriminability from the linear utility function (utility = outcome). We also compare them using 10-armed bandit tasks. We find computationally that subjective discriminability can be increased by the use of the experienced-utility function with appropriate framing for a domain, and that bigger subjective discriminability leads to more optimal decisions.

Background and Related Work

Kahneman’s utility taxonomy is useful for distinguishing multiple concepts of utility (Kahneman 2000). First, in modern economics, utility is inferred from observed choices and in turn used to explain choices. This behavioral and motivational concept of utility is called “decision utility.” Second, “experienced utility” refers to the experiences of pleasure and pain, as Bentham used it (Kahneman, Wakker, and Sarin 1997). It is the affective or hedonic impact of an obtained outcome after a choice. Kahneman distinguished experienced utility from decision utility. Recent findings in neuroscience suggest that the neural substrates of liking (experienced utility) are separate from those of wanting (decision utility) in the human brain (Berridge and Robinson 2003; Berridge and Aldridge 2006). Third, “predicted utility” is a belief about the future experienced utility of a choice before making a decision.

The role of subjective prediction in one-shot decision making under risk has been extensively examined in those outcomes are evaluated as gains (or losses), then, the frame is called gain frame (or loss frame, respectively).

In the decision-making literature (Glimcher and Rustichini 2004; Barron and Erev 2003), decisions under “risk” (when outcome probabilities of each option are explicitly described and fully known to the decision maker) are often distinguished from decisions under “uncertainty” (when outcome probabilities of each option are not explicitly described and should be learned from experiences).

We define the value of risk (VOR) for an outcome distribution as the difference between the subjective value of the outcome distribution \(X\) and that of its certainty-equivalent \(\mu_e\): VOR = \(f(X) - f(\mu_e)\) where \(f\) is the decision maker’s subjective value function. Note that the value of \(f(X)\) depends not only on \(\mu_e\) and \(\sigma_f^2\) but also on the risk attitude \((a, b, \lambda, x_{ref})\): VOR < 0 (risk aversion) if 0 < \(a\) < 1 in gain frame or \(b > 1\) in loss frame; VOR > 0 (risk seeking) if \(a > 1\) in gain frame or \(0 < b < 1\) in loss frame.
Experience-based Mode and Total-Experienced Utility

Past emotional experiences associated with a candidate option in similar situations to the current state are automatically retrieved from episodic memory and reactivated in short-term memory (Bechara et al. 1997; Niedenthal 2007). This overall reactivation, called the “experience-based mode” in our model, contributes to the motivation of selecting the option. The experience-based mode is approximated by a model-free caching reinforcement learning (RL) algorithm (Sutton and Barto 1998), which can be related to Kahneman’s moment-based approach. According to Kahneman (Kahneman 2000), “total-experienced utility” (a.k.a. “total utility”) is a statistically aggregated overall value over past experienced utilities. Total-experienced utility (or the experience-based mode) explains the role of past experiences in the computation of decision utility. 5

Discriminability

The concept of discriminability has been largely investigated under different names in a variety of areas such as psychophysical judgment and decision theory (Thurstone 1927; Holland 1975; Busemeyer and Townsend 1993), pattern classification (Duda, Hart, and Stork 2001), signal detection theory (called the “sensitivity index” or d′) (Wickens 2002) and statistical power analysis (called the “effect size”) (Cohen 1992). Discriminability can be used for characterizing the level of easiness for a task in discriminating which option is optimal with a given number of trials. Thus, as discriminability for a task becomes larger, this means that it is easier for the decision maker to tell which option is better than others in terms of average outcome.

Decisions under Uncertainty and Frames

We compare objective discriminability with subjective discriminability in two-armed bandit problems with stationary distributions of stochastic outcomes, and show that subjective discriminability can be increased by the use of the experienced-utility function with appropriate framing for a task domain.

Two-armed Bandit Tasks

Consider a two-armed bandit task in which each option k (=1, 2) is associated with a unknown normal (Gaussian) outcome distribution \( r \sim N(\mu_k, \sigma_k^2) \) (assuming \( \mu_1 > \mu_2 \)). Note that, in this paper, option 1 always denotes the optimal option, whereas option 2 is suboptimal. The goal of the decision maker is to maximize the total outcome during \( N \) trials. For the simplicity of explanation, we consider a decision-making strategy in which the decision maker clearly distinguishes initial 2\( n_B \) exploratory trials from later \( N - 2n_B \) trials (assuming 2\( n_B < N \)). Also, it is assumed that during the exploratory trials, the decision maker alternatively selects one of the options; thus, after these trials, random outcomes of \( n_B \) trials for each option will be obtained.

Objective Discriminability

To define a concept of discriminability associated with the initial 2\( n_B \)-trial exploration, we focus on the trial \( t_B(= 2n_B + 1) \) immediately after 2\( n_B \) exploratory trials. On this trial the average outcome (sample mean) of \( n_B \) observed outcomes after \( n_B \) exploratory trials of each option \( k = 1, 2 \) is computed as \( \hat{\mu}_k^{t_B} = \frac{1}{n_B} \sum_{i=1}^{n_B} r_k^{(i)} \) where \( r_k^{(i)} \) is the \( i \)th sampled outcome of option \( k \). Also, sample means \( \hat{\mu}_k^{t_B} \) follow normal distributions: \( \hat{\mu}_k^{t_B} \sim N(\mu_k, \sigma_k/\sqrt{n_B}) \) for each \( k \). Denote the option selected on trial \( t_B \) by \( a_{t_B} \). Assuming that the decision maker selects the option with higher average objective outcome, the expected frequency rate of choosing option 1 over option 2 on trial \( t_B \) in a large number of tasks is \( \text{Pr}_{\text{obj}}(a_{t_B} = 1) = \text{Pr}(\hat{\mu}_1^{t_B} > \hat{\mu}_2^{t_B}) = \text{Pr}(\hat{\mu}_1^{t_B} - \hat{\mu}_2^{t_B} > 0) = \text{Pr}(y > 0) \) where \( y = \hat{\mu}_1^{t_B} - \hat{\mu}_2^{t_B} \). Since \( \hat{\mu}_1^{t_B} \) and \( \hat{\mu}_2^{t_B} \) are normal variables, \( y \) is also a normal variable following \( y \sim N(\mu_1 - \mu_2, (\sigma_1^2 + \sigma_2^2)/n_B) \). Now the standard normal variable \( z = \frac{y - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}/n_B} \sim N(0, 1) \) whose cumulative distribution function (cdf) is \( \Phi(x) = \frac{1}{2} \left( 1 + erf \left( \frac{x}{\sqrt{2}} \right) \right) \) leads to \( \text{Pr}(y > 0) = \text{Pr}(z > -d_B) = 1 - \Phi(-d_B) = \Phi(d_B) \) where \( d_B = \sqrt{\frac{(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}/n_B} \).

Defining the objective discriminability (called objective d-prime) \( d_{\text{obj}} \) \( d_{\text{obj}} \triangleq \sqrt{\frac{(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2}/n_B} \) and thus, \( \text{Pr}_{\text{obj}}(a_{t_B} = 1) = \Phi(\sqrt{n_B}d_{\text{obj}}) \). Note that \( d_{\text{obj}} \) depends only on the statistics of objective outcome distributions given in the problem and that as \( d_{\text{obj}} \) of the underlying domain increases, the objective decision maker’s expected frequency rate of choosing option 1 over option 2 after 2\( n_B \) exploratory trials becomes close to 1.

Subjective Discriminability

Now consider what happens to the discriminability when the decision maker employs the subjective value (experienced-utility) function. Given the experienced-utility function \( f_{EU} \), the average subjective value of option \( k \) after \( n_B \) exploratory trials is the sample mean of \( n_B \) subjective values, \( \Delta_{\text{subj},k}^{t_B} \triangleq \frac{1}{n_B} \sum_{i=1}^{n_B} v_k^{(i)} \) where \( v_k^{(i)} = f_{EU}(r_k^{(i)}) \). When we approximate the distributions of the subjective-value sample means \( \mu_{\text{subj},k}^{t_B} \) by normal distributions: \( \mu_{\text{subj},k}^{t_B} \sim N(\mu_{\text{subj},k}, (\sigma_{\text{subj},k}/\sqrt{n_B})^2) \) for option \( k = 1, 2 \).

Assuming that the decision maker selects the option with higher average subjective value, the probability (i.e., expected frequency rate) of choosing option 1 over option 2 on trial \( t_B \) in a large number of tasks is \( \text{Pr}_{\text{subj}}(a_{t_B} = 1) = \Phi(\sqrt{n_B}d_{\text{subj}}) \) where the subjective discriminability (called subjective d-prime) \( d_{\text{subj}} \) \( d_{\text{subj}} \triangleq \sqrt{\frac{(\mu_{\text{subj},1} - \mu_{\text{subj},2})}{(\sigma_{\text{subj},1}^2 + \sigma_{\text{subj},2}^2)/n_B}} \). Note that \( d_{\text{subj}} \) depends not only on the underlying outcome distributions, but also on the experienced-utility function whose
Probability

Figure 4: Discriminabilities vs. reference point, showing how the reference point selection influences the decision maker’s expected frequency rate of choosing option 1 over option 2 after 2nB exploratory trials becomes close to 1.

Comparison between Objective and Subjective Discriminabilities

The decision maker’s expected frequency rate of choosing option 1 over option 2 after nB trials of each option depends on their discriminability (d'\(_{obj}\) or d'\(_{subj}\)): Pr\(_{obj}\)(a\(_t\) = 1) = \(\Phi(\sqrt{n_B} d'_{obj})\) or Pr\(_{subj}\)(a\(_t\) = 1) = \(\Phi(\sqrt{n_B} d'_{subj})\). Therefore, if subjective discriminability d'\(_{subj}\) is greater than objective discriminability d'\(_{obj}\), for a decision maker with appropriate shape and reference point of the experienced-utility function, subjective decision making can provide a better overall performance due to a higher probability of choosing option 1 over option 2 on remaining trials. In other words, to reach a pre-specified probability of selecting the optimal option, a subjective decision making with a larger d'\(_{subj}\) should require fewer exploratory trials than objective decision making with a smaller d'\(_{obj}\). Note that d'\(_{obj}\) relies only on the true means and standard deviations of underlying outcome distributions (\(\mu_1, \mu_2, \sigma_1, \sigma_2\)), whereas d'\(_{subj}\) (or \(\mu_{subj,1}, \mu_{subj,2}, \sigma_{subj,1}, \sigma_{subj,2}\)) depends on subjective value function shape parameters and reference point a, b, \(x_{ref}\) as well as \(\mu_1, \mu_2, \sigma_1, \sigma_2\).

Given a representative subjective value function (experienced-utility function) shape and reference point selection for each outcome, Figure 3 shows how the objective and subjective discriminabilities can be defined if the underlying outcome distributions were known. Here, subjective value distributions are the transformation of objective outcome distributions through the function. Here we use Monte Carlo simulations to estimate the true means (\(\mu_{subj,k}\)) and standard deviations (\(\sigma_{subj,k}\)) of the subjective value distributions (\(v_k = f_{EU}(r_k)\) for \(k = 1, 2\)) obtained by shifting the original objective outcome distributions (\(r_k \sim N(\mu_k, \sigma_k^2)\)) through the subjective value function \(f_{EU}(\cdot)\).

The Influence of Domain and Framing on the Subjective Discriminability

Subplots (a), (b) and (c) in Figure 4 show the simulation results on how the reference point selection (framing) influences subjective discriminability on different domains (Domain 1, Domain 2, and a domain where two options have equal uncertainty in outcomes) for a decision maker employing a subjective value function (experienced-utility (EU)) function with shape parameters \(a = 0.8, b = 0.5, \lambda = 2.5\).  

\footnote{From our sensitivity tests of parameters, we can obtain the same characteristics of domain-frame interaction effects when 0 <
Note that these three domains can represent all possible cases of stationary gaussian outcome distributions in two-armed bandit problems. It should be noted that $d'_{obj}$ significantly changes as the reference point selection changes, while $d'_{obj}$ does not depend on the reference point.

On Domain 1 (Figure 4 (a)), option 1 ($\mu_1 = 5$ and $\sigma_1 = 5$) is optimal with less uncertainty, while option 2 ($\mu_1 = -5$ and $\sigma_1 = 10$) is suboptimal with more uncertainty. In this domain, the gain frame ($-10 < x_{ref} < 2.5$) leads to an increased subjective discriminability ($d'_{subj} > d'_{obj}$), whereas the loss frame ($x_{ref} > 2.5$) leads to a decreased subjective discriminability ($d'_{subj} < d'_{obj}$). 7 According to the characteristic of the experienced-utility function, the decision maker’s subjective experience in the gain frame (e.g., $x_{ref} = -5$, green line) would mainly elicit the uncertainty-averse and loss-averse attitude ($0 < a < 1$, $\lambda > 1$) tending to prefer the option 1 that generates more certain gains and avoid the option 2 that generates big losses very often. The loss frame (e.g., $x_{ref} = 10$, green line) would mainly bring out the uncertainty-seeking and loss-averse attitude ($0 < b < 1$, $\lambda > 1$) tending to prefer the option 2 that generates gains sometimes and avoid the option 1 that generates more certain losses. People tend to avoid certain losses more than uncertain losses. Yet, the framing does not influence $d'_{obj}$.

On Domain 2 (Figure 4 (b)), option 1 ($\mu_1 = 5$, $\sigma_1 = 10$) is optimal with more uncertainty, while option 2 ($\mu_1 = 5$, $\sigma_1 = 5$) is suboptimal with less uncertainty. In this domain, the loss frame ($-4 < x_{ref} < 10$) leads to an increased subjective discriminability ($d'_{subj} > d'_{obj}$), whereas the gain frame ($x_{ref} < -4$) leads to a decreased subjective discriminability ($d'_{subj} < d'_{obj}$). Note that the gain frame ($x_{ref} = -10$, green line) would elicit the uncertainty-averse and loss-averse attitude ($0 < a < 1$, $\lambda > 1$) tending to prefer option 2 that generates more certain gains and avoid option 1 that generates losses sometimes. The loss frame ($x_{ref} = 5$, green line) would bring out the uncertainty-seeking and loss-averse attitude ($0 < b < 1$, $\lambda > 1$) tending to avoid the option 2 that generates more certain losses and prefer option 1 that generates big gains very often.

On Domain with equal uncertainty (Figure 4 (c)), option 1 ($\mu_1 = 5$ and $\sigma_1 = 5$) is optimal, and option 2 ($\mu_1 = -5$ and $\sigma_1 = 5$) is suboptimal. In this domain, the neutral frame (5 < $x_{ref} < 5$) leads to an increased subjective discriminability $a < 1$, $0 < b < 1$ and $\lambda > 1$. Yet, other conditions like $a > 1$ (risk-seeking when faced with likely gains), $b > 1$ (risk-averse when faced with likely losses), and/or $0 < \lambda < 1$ (loss-seeking) can bring different risk attitudes. Subjective value function parameters (shape and reference point) determine risk attitudes and change subjective discriminability.

### Figure 5: The influence of outcome uncertainties on discriminabilities for each domain.

(a): Domain 1 with $\sigma_2 = 2\sigma_1$, (b): Domain 2 with $\sigma_2 = 0.5\sigma_1$

The influence of outcome uncertainties on discriminabilities

Figure 5 illustrates how the outcome uncertainties of two options ($\sigma_1$ and $\sigma_2$) influence discriminabilities when the decision maker employs different subjective value functions. First, subplot (a) shows simulation results on Domain 1 where $\mu_1 - \mu_2 = 10$ (fixed), $\sigma_1$ is varying from 1 to 5, and $\sigma_2 = 2\sigma_1$. Second, subplot (b) shows simulation results on Domain 2 where $\mu_1 - \mu_2 = 10$ (fixed), $\sigma_1$ is varying from 2 to 10, and $\sigma_2 = 0.5\sigma_1$. On both domains the subjective discriminability is reliably greater than the objective discriminability when the levels of outcome uncertainties of each option are not very large.

### Objective and Subjective Decision Rules for Exploitative Trials

Here we introduce objective and subjective versions of greedy selection rule using objective outcomes and subjective values, respectively, but the same can be extended to other selection rules (e.g., softmax).

#### Greedy selection based on objective outcomes

After an initial $2n_B$ exploratory trials, the decision maker employs the greedy selection rule based on objective outcomes. The mean of sampled outcomes of option $k = 1, 2$
Subjective

0.0974

0.06

0.1984

0.5450

Frequency	
  rate
  (optimal	
  option)

0.5

0.5

0.6

0.6

0.7

0.7

0.8

0.8

0.9

0.9

1

1

0

0

10

10

20

20

30

30

40

40

50

50

(a) Domain 1

(b) Domain 2

Frequency rate (optimal option)

Frequency rate (optimal option)

Subj

Obj

Sampling

VPI

Table 1: 10-armed bandit experiments: softmax decision rules based on objective means or subjective means each is denoted as \( \hat{\mu}_k^t \triangleq \frac{1}{n_k^t} \sum_{i=1}^{n_k^t} r_k^{(i)} \) where \( n_k^t \) is the number of sampled outcomes of option \( k \) before trial \( t \). If \( \hat{\mu}_1^t \) is greater or lower than \( \hat{\mu}_2^t \), the decision maker selects option 1 or 2, respectively. Otherwise, they take a random action. With this rule, the expected frequency rate of selecting the optimal option on trial \( t \) is \( \Pr(\hat{\mu}_1^t > \hat{\mu}_2^t) \).

Greedy selection based on subjective values

The mean of sampled subjective values of option \( k \) (\( = 1, 2 \)) is denoted as \( \hat{\mu}_{subj}^t(k) \triangleq \frac{1}{n_k^t} \sum_{i=1}^{n_k^t} v_k^{(i)} \) where \( v_k^{(i)} = \exp\{\beta q_i^t\} \) and \( n_k^t \) is the number of sampled outcomes of option \( k \) (\( = 1, 2 \)) before trial \( t \). Also, \( \beta_{subj}^t \) denotes the variance estimate of subjective values of option \( k \) on trial \( t \). After an initial \( 2n_B \) exploratory trials, if \( \hat{\mu}_{subj,2} \) is greater or lower than \( \hat{\mu}_{subj,1}^t \), the decision maker selects option 1 or 2, respectively. Otherwise, they take a random action. Here the expected frequency rate of selecting the optimal option on trial \( t \) is \( \Pr(\hat{\mu}_{subj,1}^t > \hat{\mu}_{subj,2}^t) \).

Experiments

We compare different decision strategies such as subjective value-based greedy selection, objective outcome-based greedy selection, action value sampling (probability matching), and myopic value of perfect information (VPI) (Darden, Friedman, and Russell 1998) on Domain 1 and Domain 2. We performed 500 tasks on each domain and rule. Figure 6 shows the actual frequency rate of selecting the optimal option on trial \( t \). In both simulations (Domain 1 and Domain 2), each strategy had an initial 10 exploratory trials (\( n_B = 5 \) trials for each option). For action value sampling and myopic VPI, exploratory trials were used to initialize mean and variance priors for learning in later trials. For subjective value-based greedy selection, the reference point on each domain was set to the average outcome of the more uncertain option (gain framing on Domain 1 and loss framing on Domain 2) to obtain an increased subjective discriminability as described in the previous section. On each domain the subjective value-based greedy selection rule obtains the greatest frequency rate of selecting the optimal option over trials; and thus, the greatest total outcome.

To see if subjective experience-based learning can win against objective outcome-based learning in more generalized settings, we also performed multi-armed bandit experiments with a different decision rule. Here we compared the softmax decision rules \( \Pr(\text{option } = i) = \exp[\beta q_i^t]/\sum_{k=1}^{K}\exp[\beta q_k^t] \) based on objective means \( (q_i^t = \mu_i^t) \) or subjective means \( (q_i^t = \mu_{subj,i}^t) \) each on the 10-armed bandit domain (\( K=10 \) and 500 trials in each task) where \( \mu_i - \mu_{i+1} = 1 \) (\( i = 1, \ldots, 9 \)) and \( \sigma_i = 1 \) (\( i = 1, \ldots, 10 \)). Also, the reference point for evaluating subjective values dynamically changed over trials, setting it to the mean of the observed top two average outcomes. With \( \beta = \beta_0 t \), we report the best \( \beta_0 \) constant over 500 tasks for each case in Table 1. The results confirm that the subjective learner beats the objective learner in terms of mean loss per trial.

Discussion and Conclusion

Iowa and Shiv et al’s experiments were performed in the face of likely gains (the gain frame). In Figure 4, Iowa task corresponds to Domain 1, and has greater subjective discriminability in the gain frame than objective discriminability. However, Shiv’s task corresponds to Domain 2, and has lower subjective discriminability in the gain frame than objective discriminability.

Myopic value of perfect information (VPI) can be viewed as a sort of exploration bonus provided to outcome uncertainty under the belief that the new information gathered from the option with more uncertainty would be more likely to change the future decision strategy than that from other options with less uncertainty; thus, VPI-based learning explore the option with more uncertain outcomes more often. In contrast, some well-known economic models of choice such as the Markowitz-Tobin (MT) portfolio selection model make a trade-off between mean \( (\mu) \) and outcome variance \( (\sigma^2) \) in computing the expected utility of an option (Real 1991): expected utility \( = \mu - a\sigma^2 \) where \( a \) (\( > 0 \)) is the risk-aversion coefficient; thus, as the outcome uncertainty of an option becomes greater, choice preference for that option becomes lower. Yet, subjective learning shows different uncertainty attitudes relying on the frame chosen by the decision maker (reference point) and the shape of subjective value function parameters. With the representative function shape in Figure 2, subjective learning tends to avoid the option with more uncertain outcomes in the gain frame but prefer such an option in the loss frame.

Our contribution functions in a way that can be used computationally by AI researchers who want their systems to exhibit more of the behaviors that people exhibit.
Acknowledgements
We appreciate the generous support of MIT Media Lab consortium sponsors for this work.

References


