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; Yechiam et al. 2005) and Shiv *et al.*'s experiment (Shiv et al. 2005). Essentially both experiments can be thought of as two-armed bandit tasks involving choices between two options with different uncertain outcome distributions. In these tasks, the decision maker should regulate the balance between exploration (choices to find new information) and exploitation (choices to maximize outcome with current information) in order to maximize the overall outcome for total trials (Sutton and Barto 1998).

In the Iowa gambling task, choices are made between one option with higher mean and less uncertain outcomes (option 1) vs. the other option with lower mean and more uncertain outcomes (option 2) (e.g., Domain 1 in Figure 1) has shown that normal people are good at quickly selecting the long-run advantageous option in this type of task, whereas patients with 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 the Iowa gambling task, it should be noted that the optimal option involves *safer gain outcomes*, whereas the suboptimal option involves riskier outcomes with a long-run expected loss.

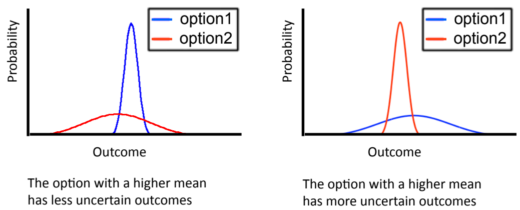
In Shiv *et al.*'s experiment (Shiv et al. 2005), choices are made between an option with higher mean and more uncertain outcomes (option 1) vs. an option with lower mean and less uncertain outcomes (option 2) (e.g., Domain 2 in Figure 1); this experiment has presented the harmful side of subjective emotional learning in terms of optimal decision behavior.[1] Shiv et al.'s 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, expected return = \$1.25) and the safer suboptimal option (no investment, \$1 gain for sure choice, expected return = \$1). People with no diagnoses 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, considering the interaction of framing and task domain. Our work contributes a novel unified framework that explains the Iowa experiment, Shiv et al's experiment, and more.

Both Iowa and Shiv et al's experiments illustrate that normal people tend to have uncertainty-averse and loss-averse attitudes 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.

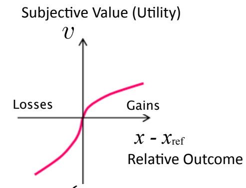
We propose that subjective experience-based learning depends mainly on the distribution of options in the task domain (mean and uncertainty) and where the outcomes lie relative to the learner's perceived reference point – the gain/loss framing of the decision. If the decision maker's own reference point for evaluating outcomes is smaller than most sampled outcomes, then those outcomes are evaluated as gains and the frame is called a gain frame. If the decision

[1]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.



(a) Domain 1 (b) Domain 2

Figure 1: Domains under uncertainty



$$v = f_{\mathrm{PT}}(x - x_{\mathrm{ref}}) = \begin{cases} (x - x_{\mathrm{ref}})^a, & x - x_{\mathrm{ref}} \geq 0 \\ -\lambda(x_{\mathrm{ref}} - x)^b, & x - x_{\mathrm{ref}} < 0 \end{cases}$$
$$0 < a < 1, \quad 0 < b < 1, \quad \lambda > 1$$

Figure 2: Prospect Theory (PT) Subjective Value Function

maker's reference point is larger than most outcomes, they are perceived as losses and the frame is a loss frame.

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 ease in figuring out which option is optimal with fewer trials. Thus, greater discriminability is a key factor in regulating the trade-off between exploration and exploitation, allowing the learner to more quickly detect the optimal decision.

To model subjective experience, we propose a new "experienced-utility function," which adds parameters to a prospect theory (PT)-based subjective value function (Figure 2) (Kahneman and Tversky 1984; Kahneman 2003).

Using two-armed bandit task simulations, we compare subjective discriminability from our experienced-utility function (utility = PT-based subjective value) with objective discriminability from the linear utility function (utility = outcome). We also run comparisons with more complex 10-armed bandit decisions. Importantly, we find that subjective discriminability can be increased by the use of the experienced-utility function with appropriate framing, 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 is used to explain choices. This behavioral and motivational concept of utility is called "decision utility." Another type of utility, "experienced utility" refers to the experiences of pleasure and pain, as Bentham used it (Kahneman, Wakker, and Sarin 1997); this 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[2] has been extensively examined in prospect theory (PT) (Kahneman 2003; Tversky and Kahneman 1992; Kahneman and Tversky 1979; 1984). PT employs a subjective value function (Figure 2) called the "predicted-utility function" by which the decision maker's risk attitudes and framing in prediction and decision can be described. Also, decisions under risk are assumed to rely on predicted utilities (i.e., decision utility = predicted utility). However, the role of subjective experience in making decisions under uncertainty is not yet well studied. In decision making under uncertainty, the overall experience of previous trials on the decision-maker has a critical impact on future decisions (i.e., decision utility $\simeq$ total-experienced utility).

Prospect Theory and Subjective Value Function

The PT subjective value function in Figure 2 has three essential characteristics: First, gains and losses are defined relative to a reference point, x_{ref}, which dictates the placement of the vertical axis shown in Figure 2. If an expected outcome x is greater or smaller than a reference point x_{ref}, then the outcome $x - x_{\mathrm{ref}}$ is viewed as a gain or a loss, respectively. The reference point may depend on framing (the way the task is designed and described) and the decision maker's expected outcome. If expecting a high positive value outcome, then a low positive outcome might be perceived in a loss frame since it lies "to the left" of the value that was expected. Second, the function has *diminishing sensitivity*: it is concave in the area of gains ($0 < a < 1$, denoting *risk-averse attitude when faced with likely gains*) and convex in the area of losses ($0 < b < 1$, denoting *risk-seeking attitude when faced with likely losses*). Third, the function is steeper in the area of losses ($\lambda > 1$, denoting *loss aversion*). [3] Note that, while PT uses the subjective value function to model their "predicted-utility function", we propose and test a PT-based parameterized subjective value function to model the "experienced-utility function." We assume that the two functions are independent of and separate from each other.

[2]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).

[3]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 (μ_x): VOR = $f(X) - f(\mu_x)$ where f is the decision maker's subjective value function. Note that the value of $f(X)$ depends not only on μ_x and σ_x^2 but also on the risk attitude (a,b,λ, 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. [4]

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.

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 $r \sim \mathrm{N}(\mu_k, \sigma_k^2)$ (assuming $\mu_1 > \mu_2$). For clarity 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 simplicity of explanation, we consider a decision-making strategy in which the decision maker clearly distinguishes the initial $2n_B$ exploratory trials from the later $N - 2n_B$ trials (assuming $2n_B < N$). Also, it is assumed that during the exploratory trials, the decision maker selects from both options; thus, after these trials, random outcomes of n_B trials for each option are obtained.

[4] Total-experienced utility could be also associated with "action value" in model-free RL and "anticipatory emotion" in the decision making literature (Bechara and Damasio 2005; Cohen, Pham, and Andrade 2006; Loewenstein and Lerner 2003; Pham 2007).

Objective Discriminability

To define a concept of discriminability associated with the initial $2n_B$-trial exploration, we focus on the trial $t_B (= 2n_B + 1)$ immediately after $2n_B$ 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 $\hat{\mu}_k^{t_B} \triangleq (1/n_B) \sum_{i=1}^{n_B} r_k^{(i)}$ where $r_k^{(i)}$ is the ith sampled outcome of option k. Also, sample means $\hat{\mu}_k^{t_B}$ follow normal distributions: $\hat{\mu}_k^{t_B} \sim \mathrm{N}(\mu_k, \left(\sigma_k/\sqrt{n_B}\right)^2)$ 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 $\Pr_{obj}(a_{t_B} = 1) = \Pr(\hat{\mu}_1^{t_B} > \hat{\mu}_2^{t_B}) = \Pr(\hat{\mu}_1^{t_B} - \hat{\mu}_2^{t_B} > 0) = \Pr(y > 0)$ where $y \triangleq \hat{\mu}_1^{t_B} - \hat{\mu}_2^{t_B}$. Since $\hat{\mu}_1^{t_B}$ and $\hat{\mu}_2^{t_B}$ are normal variables, y is also a normal variable following $y \sim \mathrm{N}(\mu_1 - \mu_2, (\sigma_1^2 + \sigma_2^2)/n_B)$. 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 $\Phi(x) = \frac{1}{2}\left(1 + erf(\frac{x}{\sqrt{2}})\right)$ leads to $\Pr(y > 0) = \Pr(z > -d_B) = 1 - \Phi(-d_B) = \Phi(d_B)$ where $d_B = \frac{\mu_1 - \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n_B}}$.

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

Subjective Discriminability

Now consider what happens to the discriminability when the decision maker employs the proposed 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, $\hat{\mu}_{subj,k}^{t_B} \triangleq (1/n_B) \sum_{i=1}^{n_B} v_k^{(i)}$ where $v_k^{(i)} = f_{EU}(r_k^{(i)})$. We approximate the distributions of the subjective-value sample means $\hat{\mu}_{subj,k}^{t_B}$ by normal distributions: $\hat{\mu}_{subj,k}^{t_B} \sim \mathrm{N}(\mu_{subj,k}, \left(\sigma_{subj,k}/\sqrt{n_B}\right)^2)$ 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 $\Pr_{subj}(a_{t_B} = 1) = \Phi(\sqrt{n_B}\, d'_{subj})$ where the *subjective* discriminability (called *subjective d-prime*) $d'_{subj} \triangleq \frac{\mu_{subj,1} - \mu_{subj,2}}{\sqrt{(\sigma_{subj,1}^2 + \sigma_{subj,2}^2)}}$. Note that d'_{subj} depends not only on the underlying outcome distributions, but also on the experienced-utility function whose shape and reference point are described by the parameters.

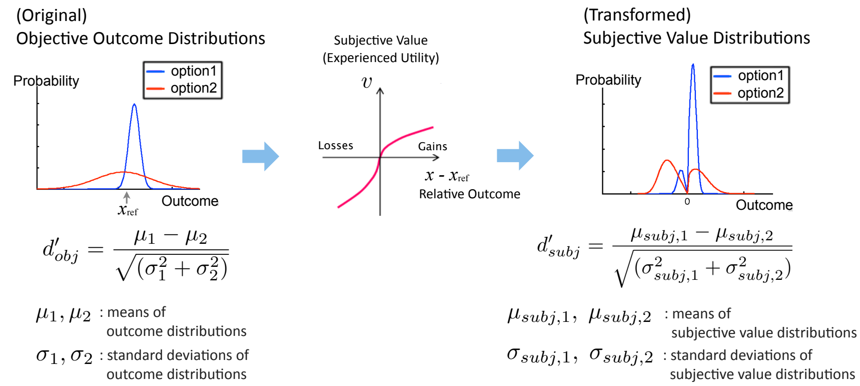


Figure 3: Objective discriminability d'_{obj} vs. Subjective discriminability d'_{subj}

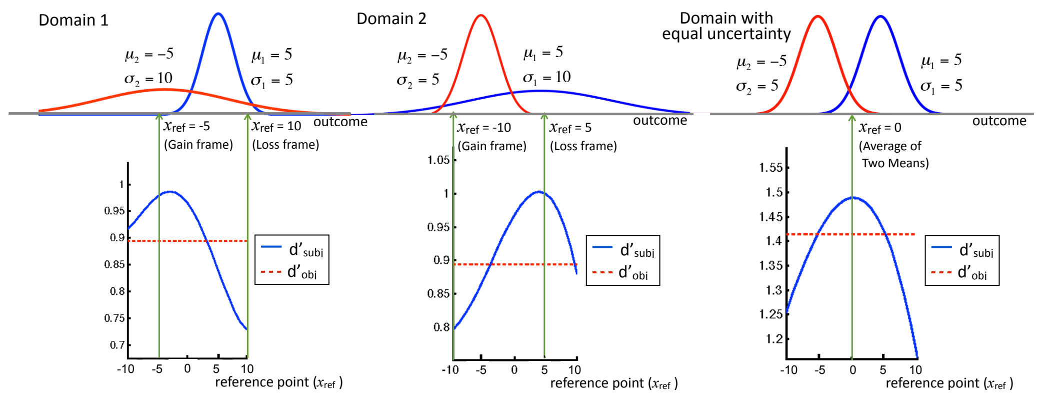


Figure 4: Discriminabilities vs. reference point, showing how the reference point selection influences d'_{subj}: (a) Domain 1; (b) Domain 2; (c) Domain with equal uncertainty. Green lines indicate example reference points to explain framing effects.

As d'_{subj} increases, the *subjective* decision maker's expected frequency rate of choosing option 1 over option 2 after $2n_B$ 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 n_B trials of each option depends on their discriminability (d'_{obj} or d'_{subj}): $\mathrm{Pr}_{obj}(a_{t_B} = 1) = \Phi(\sqrt{n_B}\, d'_{obj})$ or $\mathrm{Pr}_{subj}(a_{t_B} = 1) = \Phi(\sqrt{n_B}\, d'_{subj})$. 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, then subjective decision making can provide better overall performance due to a higher probability of choosing option 1 over option 2 on the remaining trials. In other words, to reach a pre-specified probability of selecting the optimal option, 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 $a, b, \lambda, x_{\mathrm{ref}}$ as well as $\mu_1, \mu_2, \sigma_1, \sigma_2$.

Given a representative subjective value function (experienced-utility function) shape and a reference point selection for example, Figure 3 shows how the objective and subjective discriminabilities can be defined if the underlying outcome distributions are known. 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 shaping the original *objective outcome* distributions ($r_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$) 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 $a = 0.8, b = 0.5, \lambda = 2.5$ [5].

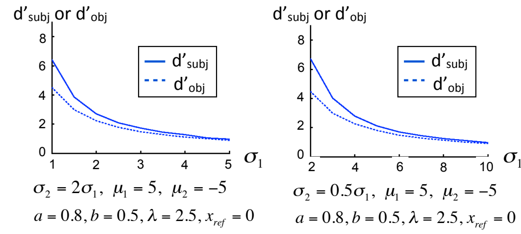
[5] From our sensitivity tests of parameters, we can obtain the same characteristics of domain-frame interaction effects when $0 < a < 1$, $0 < b < 1$ and $\lambda > 1$. Yet, other conditions like $a > 1$ (risk-seeking when faced with likely gains), $b > 1$ (risk-averse when faced with likely losses), and/or $0 < \lambda < 1$ (loss-seeking)

The three domains shown can represent all possible cases of stationary gaussian outcome distributions in two-armed bandit problems. It should be noted that d'_{subj} significantly changes as the reference point selection changes, while d'_{obj} does not depend on the reference point.

Consider Domain 1 (Figure 4 (a)), where option 1 (μ_1 = 5 and σ_1 = 5) is optimal with less uncertainty, while option 2 (μ_1 = -5 and σ_1 = 10) is suboptimal with more uncertainty. In this domain, the gain frame ($-10 < x_{\text{ref}} < 2.5$) leads to an increased subjective discriminability ($d'_{subj} > d'_{obj}$), whereas the loss frame ($x_{\text{ref}} > 2.5$) leads to a decreased subjective discriminability ($d'_{subj} < d'_{obj}$). [6] According to the characteristic of the experienced-utility function, the decision maker's subjective experience in the gain frame (e.g., $x_{\text{ref}} = -5$, green line) would mainly elicit the uncertainty-averse and loss-averse attitude ($0 < a < 1$, $\lambda > 1$) tending to prefer option 1 that generates more certain gains and avoid option 2 that often generates big losses. The loss frame (e.g., $x_{\text{ref}} = 10$, green line) would mainly bring out the uncertainty-seeking and loss-averse attitude ($0 < b < 1$, $\lambda > 1$) tending to prefer option 2 that generates gains sometimes and avoiding 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}.

Consider Domain 2 (Figure 4 (b)), where option 1 (μ_1 = 5, σ_1 = 10) is optimal with more uncertainty, while option 2 (μ_1 = -5, σ_1 = 5) is suboptimal with less uncertainty. In this domain, the loss frame ($-4 < x_{\text{ref}} < 10$) leads to an increased subjective discriminability ($d'_{subj} > d'_{obj}$), whereas the gain frame ($x_{\text{ref}} < -4$) leads to a decreased subjective discriminability ($d'_{subj} < d'_{obj}$). Note that the gain frame ($x_{\text{ref}} = -10$, 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 occasionally generates losses. The loss frame ($x_{\text{ref}} = 5$, green line) would bring out the uncertainty-seeking and loss-averse attitude ($0 < b < 1$, $\lambda > 1$) tending to prefer option 1 that can generate big gains while avoiding option 2 that generates more certain losses.

For the Domain with equal uncertainty (Figure 4 (c)), option 1 (μ_1 = 5 and σ_1 = 5) is optimal compared to option 2 (μ_1 = -5 and σ_1 = 5). In this domain, the neutral frame[7] ($-5 < x_{\text{ref}} < 5$) leads to an increased subjective discriminability ($d'_{subj} > d'_{obj}$) mainly due to loss-aversion ($\lambda > 1$, tending to avoid the option 2), whereas the gain frame ($x_{\text{ref}} < -5$) or the loss frame ($x_{\text{ref}} > 5$) leads to a



(a) Domain 1 (b) Domain 2

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 $\sigma_2 = 0.5\sigma_1$

decreased subjective discriminability ($d'_{subj} < d'_{obj}$).

In all simulations (subplots (a),(b) and (c) in Figure 4), a reference point near the mean of the average outcomes of two options leads to an increased subjective discriminability enabling more optimal decisions, regardless of the underlying outcome distributions (Domain 1, Domain 2, Domain with equal uncertainty). [8] Interestingly, when one option is more uncertain than the other option (as in Domain 1 and Domain 2), a reference point near the average outcome of the option with more uncertainty appears to maximize subjective discriminabiliity. Thus, our models shows that the decision maker is likely to have an easier time choosing optimally in the gain frame on Domain 1 and in the loss frame on Domain 2.

The influence of outcome uncertainties on discriminabilities

Figure 5 illustrates how two different outcome uncertainties (σ_1 and σ_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), σ_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), σ_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 a greedy selection rule using objective outcomes and their transformed subjective values, respectively. This approach can be extended to other selection rules, e.g., softmax.

Greedy selection based on objective outcomes

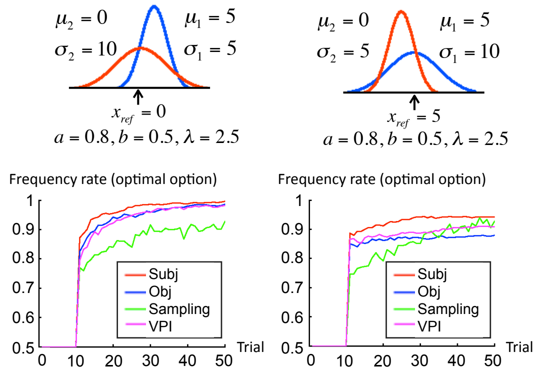
After an initial $2n_B$ exploratory trials, the decision maker employs the greedy selection rule based on objective outcomes. The mean of sampled outcomes of option k = 1, 2

can bring different risk attitudes. Subjective value function parameters (shape and reference point) determine risk attitudes and change subjective discriminability.

[6]Here we apply rough definitions on frames. For Domain 1 and Domain 2, when μ_L and μ_M indicate the average outcomes of options with less uncertainty and more uncertainty on each domain, respectively, the frame is called "gain frame" when $x_{\text{ref}} < (3\mu_L + \mu_M)/4 + \epsilon$; and "loss frame" when $x_{\text{ref}} > (3\mu_L + \mu_M)/4 + \epsilon$ for a very small positive or negative number ϵ.

[7]On the Domain with equal uncertainties, the frame is called "neutral frame" when $\mu_2 < x_{\text{ref}} < \mu_1$; "gain frame" when $x_{\text{ref}} < \mu_2$; and "loss frame" when $x_{\text{ref}} > \mu_1$.

[8]In multi-armed bandit tasks, a good reference point is the mean of the average sampled outcomes of observed best and second-best options.



(a) Domain 1 (b) Domain 2

Figure 6: The actual frequency rate of selecting the optimal option in 500 tasks on each domain for strategies (Subj: subjective greedy, Obj: objective greedy, Sampling: probability matching, VPI: myopic value of perfect information)

	Objective	Subjective
Optimal β_0	0.015	0.06
Mean of loss per trial	0.5450	0.1984
SD of loss per trial	0.5274	0.0974

Table 1: 10-armed bandit experiments: softmax decision rules based on objective means or subjective means each

is denoted as $\hat{\mu}_k^t \triangleq (1/n_k^t)\sum_{i=1}^{n_k^t} r_k^{(i)}$ 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 $\Pr_{obj}(a_t = 1) = \Pr(\hat{\mu}_1^t > \hat{\mu}_2^t)$.

Greedy selection based on subjective values

The mean of sampled subjective values of option k (= 1, 2) is denoted as $\hat{\mu}_{subj,k}^t \triangleq (1/n_k^t)\sum_{i=1}^{n_k^t} v_k^{(i)}$ where $v_k^{(i)} = f_{EU}(r_k^{(i)})$ and n_k^t is the number of sampled outcomes of option k (= 1, 2) before trial t. Also, $(\hat{\sigma}_{subj,k}^t)^2$ denotes the variance estimate of subjective values of option k on trial t. After an initial $2n_B$ exploratory trials, if $\hat{\mu}_{subj,1}^t$ is greater or lower than $\hat{\mu}_{subj,2}^t$, 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 $\Pr_{subj}(a_t = 1) = \Pr(\hat{\mu}_{subj,1}^t > \hat{\mu}_{subj,2}^t)$.

Experiments

We compare four decision strategies: subjective value-based greedy selection, objective outcome-based greedy selection, action value sampling (probability matching), and myopic value of perfect information (VPI) (Dearden, Friedman, and Russell 1998) on Domain 1 and on 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 $\Pr^t(\text{option} = i) = \exp[\beta q_i^t]/\Sigma_{l=1}^K \exp[\beta q_l^t]$ based on objective means ($q_i^t = \hat{\mu}_i^t$) or subjective means ($q_i^t = \hat{\mu}_{subj,i}^t$) each on the 10-armed bandit domain (K=10 and 500 trials in each task) where $\mu_i - \mu_{i+1} = 1$ $(i = 1, \cdots, 9)$ and $\sigma_i = 1$ $(i = 1, \cdots, 10)$. 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 β_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, the 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 explores the option with the more uncertain outcome 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 (μ) and outcome variance (σ^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 parameterized function shape we use 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

The authors appreciate the generous support of MIT Media Lab consortium sponsors for this work. The work was mainly done when Hyung-il Ahn was a doctoral candidate at the MIT Media Lab. Hyung-il Ahn also thanks IBM Research for supporting the publication.

References

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

Bechara, A., and Damasio, A. R. 2005. The somatic marker hypothesis: a neural theory of economic decision. *Games and Economic Behavior* 52(2):336–372.

Bechara, A.; Damasio, H.; Tranel, D.; and Damasio, A. 1997. Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295.

Berridge, K. C., and Aldridge, J. W. 2006. Decision utility, incentive salience, and cue-triggered 'wanting'. *To appear in Psychology of Action, 2nd Edition. Bargh, Gollwitzer, & Morsella (Eds.), Oxford University Press*.

Berridge, K., and Robinson, T. 2003. Parsing reward. *Trends in Neurosciences* 26(9).

Busemeyer, J. R., and Townsend, J. T. 1993. Decision field theory: A dynamic cognition approach to decision making in an uncertain environment. *Psychological Review* 100(3):432–459.

Cohen, J. B.; Pham, M. T.; and Andrade, E. B. 2006. The nature and role of affect in consumer behavior. *Handbook of Consumer Psychology, Curtis P. Haugtvedt, Paul Herr, and Frank Kardes (Eds.)*.

Cohen, J. 1992. Statistical power analysis. *Current Directions in Psychological Science* 1(3):98–101.

Dearden, R.; Friedman, N.; and Russell, S. 1998. Bayesian q-learning. *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)* 761–768.

Duda, R.; Hart, P.; and Stork, D. 2001. *Pattern classification*. Citeseer.

Glimcher, P. W., and Rustichini, A. 2004. Neuroeconomics: The consilience of brain and decision. *Science* 306(5695):447.

Holland, J. H. 1975. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.

Kahneman, D., and Tversky, A. 1984. Choices, values and frames. *American Psychologist* 39(4):341–350.

Kahneman, D.; Wakker, P. P.; and Sarin, R. 1997. Back to bentham? explorations of experienced utility. *The Quarterly Journal of Economics* 112(2):375–405.

Kahneman, D. 2000. Experienced utility and objective happiness: A moment-based approach. *Choices, values, and frames* 673–692.

Kahneman, D. 2003. Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review* 93(5):1449–1475.

Loewenstein, G., and Lerner, J. S. 2003. The role of affect in decision making. *Handbook of affective science* 619–642.

Naqvi, N.; Shiv, B.; and Bechara, A. 2006. The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science* 15(5):260–264.

Niedenthal, P. M. 2007. Embodying emotion. *Science* 316(5827):1002.

Pham, M. T. 2007. Emotion and rationality: A critical review and interpretation of empirical evidence. *Review of General Psychology* 11(2):155–178.

Real, L. A. 1991. Animal choice behavior and the evolution of cognitive architecture. *Science* 253(5023):980–986.

Shiv, B.; Loewenstein, G.; Bechara, A.; Damasio, H.; and Damasio, A. R. 2005. Investment behavior and the negative side of emotion. *Psychological Science* 16(6):435.

Sutton, R. S., and Barto, A. G. 1998. *Reinforcement Learning: An Introduction*. MIT Press.

Thurstone, L. L. 1927. Psychophysical analysis. *The American Journal of Psychology* 38(3):368–389.

Tversky, A., and Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty* 5(4):297–323.

Wickens, T. 2002. *Elementary signal detection theory*. Oxford University Press, USA.

Yechiam, E.; Busemeyer, J.; Stout, J.; and Bechara, A. 2005. Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science* 16(12):973.