Preference reversal or limited sampling? Maybe túngara frogs are rational after all.

Paulo Natenzon*

November 2016

Abstract

In this paper, we demonstrate that revealed preference analysis and standard welfare analysis can be applied to context-dependent choice data. We propose a new discrete choice estimation framework outside the scope of traditional random utility models, building on the individual learning model of Natenzon (2016). We identify stable (context-independent) preferences from the context-dependent choice data in Lea and Ryan (2015), and offer a new perspective in the debate about the rationality of context-dependent choice. We show that our model, the Bayesian probit, outperforms any random utility specification in goodness of fit and out of sample prediction. We conclude that our model presents a useful alternative to random utility —the current workhorse of discrete choice estimation— for applications where decision makers systematically exhibit context effects, including attraction, compromise, and phantom alternative effects.

*COMMENTS WELCOME. Department of Economics, Washington University in St. Louis, St. Louis, MO 63130, USA. E-mail: pnatenzon@wustl.edu. I am particularly indebted to John Yiran Zhu (The Wharton School, University of Pennsylvania) for bringing the dataset to my attention, and to Amanda M. Lea for many clarifications on the frog study. I also thank seminar participants at Montréal, Princeton, Columbia and The University of Chicago for useful comments and suggestions.

1 Introduction

Consider the following type of choice reversal. Alternative B is chosen in more than 50% of the choice trials in binary comparisons between A and B,

$$P(B, \{A, B\}) > 1/2$$

while alternative A is chosen in more than 50% of the choice trials in ternary comparisons among A, B and C,

$$P(A, \{A, B, C\}) > 1/2.$$

Experimenters have found many "recipes" for systematically generating such choice reversals in the lab. Famous examples include decoy effects —such as the attraction and the compromise effects— and phantom alternative effects (Huber et al. (1982), Huber and Puto (1983), Simonson (1989), Soltani, De Martino and Camerer (2012)). These examples are puzzling because they are incompatible with random utility models —the de facto current workhorse in discrete choice estimation— including logit, probit, nested logit, mixed logit etc.

The random utility framework maintains the assumption that decision makers maximize utility, while allowing the utility of each option to be a random variable. This framework can generate a rich variety of choice patterns, accommodating heterogeneity of tastes in a population, stochastic taste shocks, and hand trembling mistakes. But it cannot accommodate the type of context dependence that is observed with decoy and phantom alternative effects. Often this leads analysts to conclude that datasets in which these choice reversals can be observed are altogether incompatible with utility maximization.

Our contribution in this paper is to demonstrate that revealed preference and standard welfare analysis can be applied to context-dependent choice data. We propose a new discrete choice estimation framework outside the scope of traditional random utility models, building on the individual learning model of Natenzon (2016). Our framework is based on an "as-if" model of optimizing behavior subject to informational constraints. We maintain the assumption that agents maximize stable utility values, but relax the assumption that decision makers have perfect information about the value of each alternative when making a choice. By identifying stable preferences from datasets that may include certain kinds of choice reversals, our framework extends the scope of classic welfare analysis to applications where choice behavior is context-dependent.

To demonstrate the utility of this new framework, we identify stable (context-independent) preferences from the context-dependent choice data in Lea and Ryan (2015). The data presented in Lea and Ryan (2015) collates the mate choices of female túngara frogs in the lab. Why have we chosen data from frogs to illustrate the strength of our estimation framework? In recent years, biologists have published many accounts of irrationality in nature, borrowing models from economics, and importing many assumptions from rational choice theory. It turns out, for example, that the same experimental "recipes" that generate choice reversals for humans seem to work equally well for monkeys, frogs, birds, bees, and even slime mold (Section 11). Irrationality in nature is particularly puzzling because it is incompatible with some basic tenets of natural selection. Lea and Ryan's study is precisely concerned with preference reversals, and is used to suggest that the frogs are acting irrationally. By providing a framework in which tungara frogs are rational after all, we make sense of apparent preference reversals, thereby providing a solution to a naturalistic puzzle as well as an economic one. Our model suggests a possible mechanism for how context-dependent choice behavior may have arisen across different animal species through evolutionary pressure.

The rest of the paper is organized as follows. We present the experimental data in Section 2. We argue that the experimental design fits an identifiable pattern in experiments that generate choice reversals in humans and many other species in Section 3. We explain the difficulties in fitting the traditional random utility framework to such choice reversals in in Section 4. We describe our model in Section 5. We apply non-parametric definitions of *easier to compare*, *revealed preference* and *revealed similarity* to the data in Section 6 and derive implications for the estimation of the model. We show the conditions under which our model generates choice reversals in Section 7. We fit the model to the data in Section 8 and perform pseudo out-of-sample prediction exercises in Section 9. Our model is shown to transparently outperform the random utility framework in goodness of fit and prediction. We discuss the implications for the debate on the rationality of context-dependent choice data in Section 10. Section 11 discusses our contribution in the context of the existing literature. Section 12 concludes by suggesting three natural extensions of our framework.

2 Choice reversals in the frog data

Female túngara frogs choose mating partners based on the sound of their call. Lea and Ryan (2015) simulate three different male frog calls in the lab, which they label as target (A), competitor (B) and decoy (C). The first three rows of Table 1 show how often female frogs chose each alternative when every pairwise combination was offered: A versus B, B versus C, and A versus C. Binary comparisons are statistically significant and (stochastically) transitive: $B \succ A$, $A \succ C$, and $B \succ C$. Hence the binary choice data reveals a complete and transitive ranking of the three options: $B \succ A \succ C$.

Presented alternatives	n	A	B	C
A and B	118	.37	.63	–
B and C	90	–	.69	.31
A and C	90	.84	–	.16
A, B, and C	40	.55	.28	.17
A, B, and $\cancel{C}$	79	.61	.39	–

Table 1: Choice frequencies by female túngara frogs in the dataset of Lea and Ryan (2015). The first three rows correspond to binary choice data and support $B \succ A \succ C$ as a rational benchmark. The fourth row shows choice frequencies when all three options are available. The fifth row shows frequencies for A and B when option C was located on the ceiling, so that it was presented but unreachable. While B is more likely to be chosen than A in binary choice (first row), the opposite happens in the presence of C (last two rows).

First choice reversal. The fourth row in Table 1 shows that frogs were more likely to choose A over B when all three alternatives were offered. This contradicts the ranking $B \succ A \succ C$ obtained in binary choices.

Second choice reversal. The last row of Table 1 shows the frequencies of choice for options A and B when frogs could hear the calls A, B and C but C was a *phantom alternative*. This means that, while the three options were equidistant from the frog in the experimental chamber, A and B were placed on the floor, while option C was placed on the ceiling. Hence all three male calls could be heard, but only A and B were choosable. Comparing the first row to the last row of Table 1, we find that the presence of phantom alternative C significantly reversed the propensity of choosing A over B.

Noteworthy features of this dataset

This dataset presents a unique opportunity to demonstrate the usefulness of our new discrete choice estimation framework. We enumerate three main reasons:

1. **Simplicity:** The same three male calls A, B, C are used throughout the entire experiment. This allows us to avoid many unnecessary complications in the analysis. For example, we are able to fit the entire family of random utility models in a single, five-parameter specification. The analysis transparently shows the superiority of our model in goodness of fit and out of sample prediction.

2. **Richness:** All the pairwise combinations were separately tested, in addition to two different decoy treatments; sample sizes are large enough to allow for straightforward comparisons of choice frequencies across different treatments.

3. **Relevance:** The experimental design manipulates the options along two dimensions, copying a classic experimental design that has been found to induce context-effects in humans and many other species. But unlike many other studies, this experiment involves high-stakes decisions for the subjects. Female túngara frogs mate once or only a few times in their lifetimes.

3 Generating choice reversals in experiments

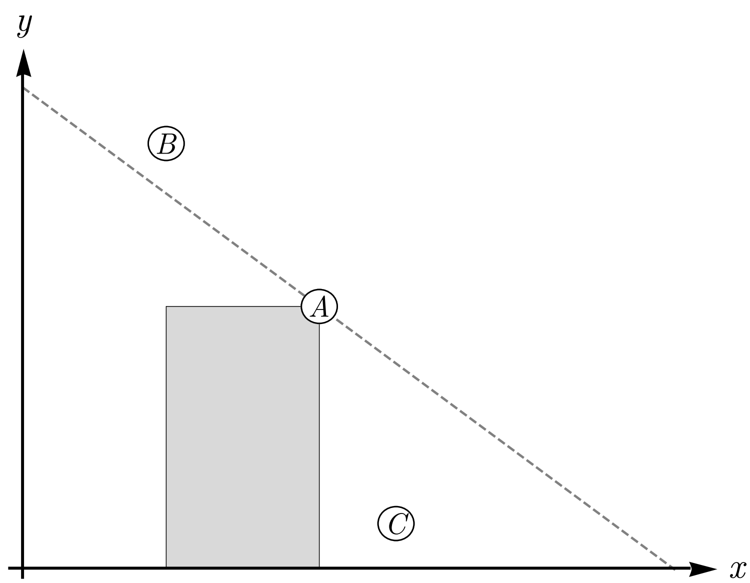

Figure 1: Male frog calls A, B and C are differentiated along two dimensions. The x-axis represents a measure of "static attractiveness", while the y-axis represents speed (see the supplemental online appendix in Lea and Ryan (2015) for details). The gray rectangle represents the location of decoy alternatives in experiments that employ asymmetric dominance.

The experimental design in Lea and Ryan (2015) can be seen to follow an identifiable pattern for experiments that induce choice reversals:

(i) Let the options A (the "target") and B (the "competitor") have different characteristics while being of comparable value, with B chosen a bit more frequently than A in binary comparisons;

(ii) Introduce an inferior alternative C (the decoy) which is **much easier to compare** to A than to B.

A particular example of (ii) above is the introduction of an asymmetrically dominated decoy: let the options be differentiated in two or more dimensions; let option A dominate the decoy in all dimensions; but let option B be worse than

C in at least one dimension. The gray rectangle in Figure 1 illustrates the location of an asymmetrically dominated decoy in the two-dimensional setting.

Figure 1 also shows the relative position of options A, B and C used in the frog experiment. The sound of male calls in túngara frogs are complex, with dozens of different measurable attributes. The experiment differentiated the options across two dimensions: the x-axis is a measure of "static attractiveness", while the y-axis measures the rate of calls per second. Both dimensions are desirable —previous studies have shown that increases in each of these dimensions lead to females choosing the option more often (see the supplemental online appendix in Lea and Ryan (2015) for details). Notice that option C is closer to A than to B along every dimension, but is not strictly dominated. As the analysis below shows, the data supports the hypothesis that C is much easier to compare to A than to B.

4 Choice reversals versus random utility

The same kind of choice reversals observed in the frog data have been found in many other settings. Every time a new experimental setup that systematically generates this type of choice reversal is discovered, it immediately becomes a puzzle. The most famous example is the attraction effect. These puzzles arise because the data cannot be rationalized by any random utility model, the current workhorse of discrete choice estimation.

Random utility models maintain the assumption of utility maximization, while incorporating a stochastic additive shock to utility. In other words, the utility of each choice alternative is written as

$$U_i = \mu_i + \varepsilon_i, \qquad i = A, B, C$$

where μ_i is the deterministic component of utility and ε_i is a random term that captures taste variation, hand trembling mistakes, and other uncontrolled factors that appear random to the econometrician. Commonly used random utility models include multinomial probit, logit, and generalizations of logit such as nested logit, mixed logit, and so on.

In this framework, the decision maker choses option A whenever the random utility U_A is greater than the random utility of every other available alternative. For every state of nature in which $[U_A > U_B \text{ and } U_A > U_C]$, we obviously also have, in particular, that $[U_A > U_B]$. Therefore, the probability of these events must satisfy:

$$\mathbb{P}[U_A > U_B] \geq \mathbb{P}[U_A > U_B \text{ and } U_A > U_C] \tag{1}$$

Hence, when agents maximize random utility, the probability of A being chosen can only decrease once alternative C is introduced.

The weak inequality in (1) holds independently of any distributional assumptions the analyst makes about the utility shocks ε_i, including allowing the shocks to be correlated. Hence, in the random utility framework the choice reversal found by comparing the first and the fourth rows of Table 1 can only arise as an anomaly and an artifact of finite sampling.

Moreover, the random utility framework does not allow phantom alternatives to have any effect on choice. When the only choosable options are A and B, the probability that A is chosen is always equal to $\mathbb{P}[U_A > U_B]$, independently of how many phantom alternatives are presented. Hence, the random utility framework constrains the choice probabilities that generate the first and last rows of Table 1 to be identical. Again, the statistically significant difference in those rows provides strong evidence that the true data generating process must lie beyond the random utility framework.

Restricting the analysis to the random utility framework, the choice behavior presented in Table 1 seems incompatible with the existence of stable, complete and transitive preferences over the choice alternatives. Often, this is interpreted as evidence of irrationality. For example, Lea and Ryan (2015) interpret the choice data as arising from a reversal of the underlying preferences:

> *Female túngara frogs reversed their preferences in the presence of an irrelevant alternative in two separate experiments and thus violate a key assumption of mate choice models derived from decision theory. (Lea and Ryan, Science, Aug 2015)*

In the next section, we present a new alternative to the random utility framework, and explain how context-dependent choice may arise from the maximization of stable utility values, under limited sampling.

5 Model: choice under limited sampling

Multiple factors contribute to uncertainty and may lead to errors in mating choices among animals. Mating choices are made in complex, dynamic environments; individuals exhibit complex traits; time spent contemplating available mate options can increase the risk of exposure to predators; organisms have limited cognitive resources; and so on. These factors lead to *limited sampling*: organisms obtain imperfect evidence about the value of each alternative before making a choice.

Building on the individual learning model of Natenzon (2016), we model the decision maker as having imperfect information about the value of the options in any particular choice situation. We assume the decision maker optimally uses the limited information available to choose the alternative with the highest expected value.

Formally, the value of the options in the population is assumed to be identically and independently distributed according to a Gaussian distribution $\mu_i \sim \mathcal{N}(m, 1/s)$. Female frogs see every choice problem, a priori, as a collection of independent draws from a population of male frogs whose utility (or "darwinian fitness") is normally distributed with mean m and variance $1/s$.

In every choice trial, the female frog obtains some information about the value of available mates based on the sound of their call. We model the information obtained about the utility of each option as noisy signals with a joint Gaussian distribution $X_i = \mu_i + \varepsilon_i$, where μ_i is the true value of alternative i and ε_i is a random perception error with $\mathbb{E}[\varepsilon_i] = 0$ and $\mathrm{Var}(\varepsilon_i) = 1/p$ for every i. We allow errors to be correlated, with $\sigma_{ij} := \mathrm{Corr}(\varepsilon_i, \varepsilon_j) \in (-1, 1)$ for all i, j.

The decision maker chooses the alternative i that maximizes $\mathbb{E}[\mu_i|X]$. Hence, the decision maker uses all the information obtained from the noisy signals to update the prior and chooses the alternative with the highest posterior mean.

Allowing Bayesian updating introduces a fundamental departure from the random utility model. For example, the multinomial probit is a classic random utility model which shares the assumption of Gaussian distributed error terms. While in the random utility framework the decision maker chooses alternative i to maximize X_i, the Bayesian probit decision maker chooses i to maximize $\mathbb{E}[\mu_i|X]$. When signals are correlated, the alternative with the highest signal need not be the same as the alternative with the highest posterior mean.

In the random utility framework, the probability that option i is ranked above option j is given by $\mathbb{P}\{X_i > X_j\}$ in every menu of alaternatives. In contrast, in the Bayesian probit the option i is ranked above j when $\mathbb{E}\{\mu_i|X\} > \mathbb{E}\{\mu_j|X\}$. The distribution of $\mathbb{E}\{\mu_i|X\}$ depends on the available signals X and may when alterantives are added to the menu. In particular

$$\mathbb{E}\{\mu_A|X_A, X_B\} \neq \mathbb{E}\{\mu_A|X_A, X_B, X_C\}$$

whenever X_C is correlated with X_A. Hence, the presence of a signal about C can influence the probability of choosing A over B.

The parameters of the prior are never identified from choice data and can without loss of generality be normalized to $\mu_i \sim \mathcal{N}(0, 1)$:

Lemma 1. *The Bayesian probit with prior $\mathcal{N}(\hat{m}, 1/\hat{s})$, utility parameters $\hat{\mu}_i \in \mathbb{R}$, precision $\hat{p} > 0$ and correlation parameters $-1 < \hat{\sigma}_{ij} < 1$ is observationally equivalent to the Bayesian probit with prior $\mathcal{N}(0, 1)$, utility $\mu_i = \sqrt{\hat{s}}(\hat{\mu}_i - \hat{m})$, precision $p = \hat{p}/\hat{s}$ and correlation $\sigma_{ij} = \hat{\sigma}_{ij}$.*

Lemma 1 has two implications for the interpretation of parameter estimates. With the normalized prior $\mathcal{N}(0, 1)$, the utility parameter μ is measured in standard deviations from the population mean. For example, $\mu_i = -0.5$ means that the darwinian fitness of frog i is half a standard deviation below the population mean. Likewise, the precision parameter $t = \hat{t}/\hat{s}$ measures the precision of the information obtained by the decision maker in units of precision in the population distribution. There is a marked difference from the random utility framework, where utility parameters have only ordinal (but not cardinal) meaning.

Let P be a stochastic choice function that describes choice probabilities in every menu of alternatives. We write $P(A, \{A, B, C\})$ for the probability that option A is chosen when A, B, C are presented. We also explicitly allow for phantom alternatives (i.e. alternatives that are presented to the decision maker but cannot be chosen). We let $P(A, \{A, B, \not{C}\})$ denote the probability that option A is chosen when A, B, C are presented to the decision maker, but only A and B are choosable. And so on.

We write $P_{p,\mu,\Sigma}$ for the stochastic choice function generated by the Bayesian probit model parameterized by the utility vector (μ_A, μ_B, μ_C) and a the covariance matrix $(1/p)\Sigma$ where

$$\Sigma = \begin{bmatrix} 1 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & 1 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & 1 \end{bmatrix}.$$

For example, $P_{p,\mu,\Sigma}(A, \{A, B, C\})$ is the probability that option A has the highest posterior mean when the decision maker receives the signals X_A, X_B, X_C. By the Bayesian updating formula, the posterior mean vector is given by

$$\begin{bmatrix} m_A \\ m_B \\ m_C \end{bmatrix} := \mathbb{E}[\mu | X_A, X_B, X_C] = \left[I + p^{-1}\Sigma\right]^{-1} \begin{bmatrix} X_A \\ X_B \\ X_C \end{bmatrix}$$

which has a joint Gaussian distribution

$$\begin{bmatrix} m_A \\ m_B \\ m_C \end{bmatrix} \sim \mathcal{N}\left(\left[I + p^{-1}\Sigma\right]^{-1} \begin{bmatrix} \mu_A \\ \mu_B \\ \mu_C \end{bmatrix}, \left[I + p^{-1}\Sigma\right]^{-1} \Sigma \left[I + p^{-1}\Sigma\right]^{-1} \right)$$

The vector of posterior means (m_A, m_B, m_C) above determines the choice probabilities from $\{A, B, C\}$ when all alternatives are choosable, and also from $\{A, B, \not{C}\}$ when C is phantom alternative. For example:

$$P_{p,\mu,\Sigma}(A, \{A, B, C\}) = \mathbb{P}[m_A > m_B \text{ and } m_A > m_C]$$
$$P_{p,\mu,\Sigma}(A, \{A, B, \not{C}\}) = \mathbb{P}[m_A > m_B]$$

In choice trials where only two alternatives i, j are presented, the decision maker receives two signals X_i, X_j, and chooses alternative i if and only if $\mathbb{E}[\mu_i - \mu_j | X_i, X_j] > 0$. By Bayesian updating, the binary choice probabilities are:

$$P_{p,\mu,\Sigma}(i, \{i, j\}) = \Phi\left(\frac{\sqrt{p}}{\sqrt{2}} \times (\mu_i - \mu_j) \times \frac{1}{\sqrt{1 - \sigma_{ij}}}\right) \tag{2}$$

where Φ is the standard Gaussian cumulative distribution function. The binary choice formula (2) shows that the ability of the decision maker to compare and correctly discriminate among the pair of options i, j increases with the overall precision of the signals p, the difference in value $|\mu_i - \mu_j|$ and the signal correlation σ_{ij}. The parameter σ_{ij} determines the comparability of alternatives i, j and can vary independently from their value. It is sometimes called the *substitutability* parameter in the discrete choice estimation literature. We refer to σ_{ij} as the *similarity* parameter following the literature in psychology (see Appendix A).

5.1 The connection to random utility models

The Bayesian probit nests two particular cases of random utility models. Suppose we restrict the domain of choice to binary comparisons, and exclude the possibility of phantom alternatives. In this case, a choice alternative in a pairwise comparison has the highest posterior mean if and only if it has the highest signal realization. Hence, the binary choice probability formula (2) is identical to the binary choice probabilities of the classic multinomial probit under the assumption that all variance parameters are equal to $1/p$.

The Bayesian probit also nests a particular case of random utility allowing for more than two alternatives. Suppose we restrict all correlation parameters to be zero $\sigma_{ij} = 0$. In this case, the similarity of the alternatives cannot play any role in making some pairs easier to compare than others. It follows in this case that an alternative has the highest posterior mean if and only if it has the highest signal realization. Thus, under the restriction $\sigma_{ij} = 0$ for all i, j our model becomes equivalent to a standard multinomial probit model with zero correlations and all variances equal to $1/p$.

5.2 Departure from random utility models

Below, we show that our model can accommodate both choice reversals in the frog data. This pattern of behavior cannot be approximated by any random utility model (Section 4). Therefore, the Bayesian probit is not nested within the random utility family.

Note that for a single, fixed menu with n choice alternatives $\{1, 2, \ldots, n\}$, the posterior mean is a random vector with a joint Gaussian distribution:

$$(m_1, m_2, \ldots, m_n) = \left[I + p^{-1}\Sigma_n\right]^{-1} (X_1, \ldots, X_n)'$$

and the decision maker picks the alternative i with the maximum realization m_i. This is analogous to a random utility model for a fixed menu. But in a random utility model, the distribution of the random value is the same in every menu. In contrast, in our model the distribution of m_i is context-dependent: it can by affected by the signals of every alternative in the menu through the correlation matrix Σ_n.

Finally, our model offers a very different interpretation of the informational asymmetry between experimental subjects and econometrician. The random utility model allows the following interpretation: each subject is perfectly informed about the value of the alternatives and maximizes utility deterministically. Utility appears to be random to the econometrician due to unobservables that cannot be controlled for. In contrast, in our model utility is random to each decision maker but not to the econometrician. Each frog sees the options in a choice trial as random draws from the same population. A frog chooses the option with the highest *expected* value under imperfect information about the value of each alternative. But from the point of view of the econometrician options A, B and C are the same across the entire experiment. The econometrician makes inference about the fixed, realized values μ_A, μ_B, μ_C by observing hundreds of choice trials using the same choice options.

6 Revealed Preference and Revealed Similarity

We provide non-parametric definitions of *easier to compare*, *revealed preference* and *revealed similarity* for any stochastic choice function P. We apply these non-parametric definitions to the choice frequency data from Table 1, and show the implications for the estimation of our parametric model $P_{\mu,\Sigma,p}$.

Definition. A pair of alternatives $\{i,j\}$ is *easier to compare* than $\{k,\ell\}$ if

$$\left|P(i,\{i,j\}) - \frac{1}{2}\right| > \left|P(k,\{k,\ell\}) - \frac{1}{2}\right|.$$

A pair is easier to compare whenever the choice frequencies in the binary comparison are more extreme (i.e., closer to zero or one). We also say that the pair is *easier to discriminate*. The first three rows in Table 1 show that $\{A,C\}$ is easier to compare than $\{B,C\}$ and, in turn, $\{B,C\}$ is easier to compare than $\{A,B\}$. In the parametric formula (2), easiness of comparison is increasing in distance in value $|\mu_i - \mu_j|$, correlation σ_{ij} and precision p. Holding fixed the precision parameter, a pair can be easier to compare because it has larger distance in value, or larger correlation, or both.

Definition. An alternative i is *revealed preferred* to j if $P(i,\{i,j\}) > 1/2$.

Write $i \succ j$ when i is revealed preferred to j. The binary choice frequencies on the first three rows of Table 1 reveal the complete and transitive ranking $B \succ A \succ C$. All the binary comparisons are easily seen to be statistically significant. From equation (2) it is clear that the parametric model accomodates the revealed preference ranking obtained from the data if and only if $\mu_B > \mu_A > \mu_C$.

Definition. $\{i,j\}$ is *revealed more similar* than $\{k,\ell\}$ if

(i) $\{i,j\}$ is easier to compare than $\{k,\ell\}$, and

(ii) $k \succcurlyeq i \succ j \succcurlyeq \ell$

When $i = k$ above, we say i is revealed more similar to j than to ℓ.

According to the first three rows of Table 1, the pair $\{A, C\}$ is revealed more similar than the pair $\{B, C\}$. In other words, C is revealed more similar to A than to B. To see this, note that (i) the pair $\{A, C\}$ is easier to discriminate than any other pair, and (ii) $B \succ A \succ C$. It follows from equation (2) that the Bayesian probit accommodates the revealed similarity relation only if $\sigma_{AC} > \sigma_{BC}$.

7 Accommodating choice reversals

Table 1 presents evidence that frogs find the pair $\{A, C\}$ easier to compare than the other pairs. The previous section concluded that, in order to accommodate the observed pattern of binary choices, we must have the correlations $\sigma_{AC} > \sigma_{BC}$. A higher degree of correlation between the signals of alternatives A and C means the signals are more informative about their difference in value:

$$X_A - X_C = (\mu_A - \mu_C) + (\varepsilon_A - \varepsilon_C)$$

where

$$\mathrm{Var}[\varepsilon_A - \varepsilon_C] = 2(1 - \sigma_{AC})/p.$$

For any fixed signal precision p, $\mathrm{Var}[\varepsilon_A - \varepsilon_C]$ goes to zero as σ_{AC} goes to one. As the correlation between ε_A and ε_C increases, these perception errors tend to "cancel out", revealing the true difference $(\mu_A - \mu_C)$ with high precision. A decision maker that observes X_A and X_C with a small precision p and high correlation σ_{AC} will learn very little about their individual values, but a great deal about the difference in values. This can be paraphrased as "I have no idea how good A and C are, but I am pretty sure that A is better than C."

When X_A and X_C are highly correlated, it is optimal *not* to choose B in some cases where X_B has the highest signal realization. For example, consider an extreme case where we fix $p > 0$ to be arbitrarily small, fix $\sigma_{BC} = \sigma_{AB} = 0$ and consider the limit as $\sigma_{AC} \to 1$. Before hearing the call of the male frogs, a female frog sees the joint distribution of the value of any three options A, B and C as exchangeable (invariant to permutations of the alternatives). Every one of the $3! = 6$ strict rankings $i \succ j \succ k$ among A, B and C are equally likely ex-ante.

Observing a signal realization for X_A, X_B, X_C with the correlations and precision above, a female will learn that either $A \succ C$ or $C \succ A$, and nothing else. Given the low precision, the signal X_B is very uninformative about the value of B, and the posterior m_B will stay very close to the mean of the population. On the other hand, the high, positive correlation between X_A and X_C make the distribution of posteriors m_A and m_C negatively correlated: learning that $A \succ C$ is good news for A, and bad news for C. Learning that $C \succ A$ is bad news for A and good news for C. In either case, the posterior m_B is very likely to end up in the middle between m_A and m_C. In other words, alternative B suffers a disadvantage when A and C are easier to compare: it needs have to obtain a very high, unlikely signal realization X_B for the posterior m_B to move enough from the mean in order to escape the "trap" created by the high value of σ_{AC}.

This intuition is made precise by analyzing the closed-form choice probabilities obtained by Natenzon (2016) for the case where $p \to 0$. The formula implies that, for any fixed values of μ_A, μ_B, μ_C, there exists a $\bar{p} > 0$ such that for every precision p in $[0, \bar{p})$, alternative A is chosen more often than B if and only if $\sigma_{AC} > \sigma_{BC}$. Hence, the presence of a decoy C that is more correlated with A than with B may induce a choice frequency reversal if B was originally chosen more often than A when C was not presented. The estimates obtained in the next Section confirm this mechanism allows our model to fit both choice reversals in the data.

8 Estimates of context-independent preferences

The choice frequencies presented in Table 1 have six degrees of freedom. The Bayesian probit model has seven parameters: μ_A, μ_B, μ_C are the utilities of each choice option, $\sigma_{AB}, \sigma_{AC}, \sigma_{BC}$ reflect the comparability of each pair of options, and p is the precision of the signals. We fit the model by maximum likelihood, imposing the following linear restrictions on the parameter space:

$$\mu_A + \mu_B + \mu_C = 0, \quad \sigma_{AC} + \sigma_{AB} = 1, \text{ and } \sigma_{BC} = 0 \tag{3}$$

The restricted version of the model has four degrees of freedom. Given Lemma 1, the restriction $\mu_A + \mu_B + \mu_C = 0$ amounts to assuming that the average utility among the three options used in the experiment matches the average in the population. This reduces the degrees of freedom of the utility function to two. Restrictions of this type are ubiquitous in applications. For example, the econometrician estimates a single risk aversion parameter that determines the utility of every option, rather than an individual parameter for each option. The remaining restrictions reduce the dimensionality of the space of correlation parameters to one. Similar restrictions are typically imposed in applications of the probit, see for example Hausman and Wise (1978). Relaxing the constraints in (3) can only improve the fit of the model (while possibly reducing the precision of the estimates), and does not change any of the qualitative results.

Parameter	Estimate	Std. Dev.
p	0.0905	0.0652
μ_A	-0.1931	0.5239
μ_B	1.2809	0.6001
μ_C	-1.0878	
σ_{AB}	0.0484	0.0999
σ_{AC}	0.9516	
σ_{BC}	0.0000	
N trials	417	
–Loglike	-266.27	

Table 2: Point estimates for utility μ_i, similarity σ_{ij} and information precision p in the Bayesian probit model imposing the additional restrictions in equation (3).

The parameter estimates of Table 2 tell a rational story of utility maximization under limited sampling. The three alternatives can be ranked in a complete and transitive manner according to value. Alternative B is the best alternative, with $\mu_B = 1.2809$, which means B is more than one standard deviation above the population mean in terms of 'darwinian fitness' (frog B is better than 90% of the population). Frog A comes in second place, lying 0.1931 standard deviation below the population mean (better than 42% of the population). Finally, frog C is the worst option, being more than one a standard deviation below the population mean (better than 14% of the population).

Given this rational benchmark, we interpret every instance in which an inferior option was chosen in the dataset as a *mistake* resulting from limited sampling. The precision parameter $p = 0.0905$ gives a measure of the information limitation. The estimated correlation parameters give a measure of the pairwise similarity or comparability of the options. With the similarity parameter σ_{BC} fixed to zero, alternatives A and C are estimated to be the most similar, with correlation $\sigma_{AC} = 0.9516$, followed by alternatives A and B with estimated correlation $\sigma_{AB} = 0.0484$. The next table shows these estimates accommodate both choice reversals observed in the experiment, and a much better fit than any random utility model.

8.1 Comparison to random utility

Table 3 compares the choice frequencies in the data to the estimated choice probabilities for the Bayesian probit (BP) and for the random utility model (RUM). Estimates were obtained by maximum likelihood. We estimated a five-parameter model of the entire RUM family (this is possible due to the simplicity of the data and to Lemma 2 below). Hence the fit of the RUM model in Table 3 is, by definition, superior to the fit of any of the special cases of RUM such as logit, probit, nested logit, mixed logit, and so on.

	data			BP			RUM		
Menu	A	B	C	A	B	C	A	B	C
$\{A, B\}$	.37	.63	–	.37	.63	–	.48	.52	–
$\{B, C\}$	–	.69	.31	–	.69	.31	–	.69	.31
$\{A, C\}$	.84	–	.16	.81	–	.19	.83	–	.17
$\{A, B, C\}$	.55	.28	.17	.59	.30	.11	.48	.35	.17
$\{A, B, \not{C}\}$	.61	.39	–	.63	.37	–	.48	.52	–

Table 3: Comparison of relative frequencies of choice in the original data (left), estimated choice probabilities for the Bayesian probit (middle) and estimated choice probabilities for the random utility model (right).

Table 3 illustrates the difficulty of RUM models to capture the choice frequency reversals in the data. First, the RUM model restricts the probability of choosing A from $\{A, B\}$ to be the same independently of the presence of the phantom alternative C. Thus, the RUM model restricts the first row and the last row of Table 3 to be identical. Moreover, in the RUM model the probability of choosing A from $\{A, B\}$ can only increase when the decoy alternative C is added to the menu. Thus, the RUM model restricts the probability of choosing A in the first row of Table 3 to be at least as large as the probability of choosing A in the fourth row of Table 3.

In the random utility model (RUM), the utility of each choice alternative is a random variable

$$\begin{aligned} U_A &= u_A + \varepsilon_A \\ U_B &= u_B + \varepsilon_B \\ U_C &= u_C + \varepsilon_C \end{aligned} \tag{4}$$

where u_i is the deterministic component of the utility of alternative i and ε_i is a stochastic taste shock for $i = A, B, C$. Under the additional assumption that the stochastic taste shocks are joint normally distributed, we obtain the multinomial probit model; if the shocks are iid Gumbel distributed, we obtain the logit model; assuming shocks have a joint Generalized Extreme Value distribution, we obtain generalizations of logit such as nested logit, cross nested logit, and so on (see any standard textbook, e.g. Train (2009)).

Instead of separately testing the fit of a random utility model under different assumptions about the distribution of taste shocks, we will use the following lemma to find a single best fit among the entire set of random utility models:

Lemma 2 (Block and Marschak, 1960)**.** *Every RUM is equivalent to a probability measure over the $n!$ strict rankings over the n choice alternatives.*

Proof. See Block and Marschak (1960), Theorem 3.1. □

Let $p_{ABC} \in [0, 1]$ denote the probability of the strict ranking $A \succ B \succ C$ and analogously denote the probability of every other strict ranking over the three alternatives. Since there is a total of $3! = 6$ such rankings, every RUM can be described by the five parameters $p_{ABC}, p_{ACB}, p_{BAC}, p_{BCA}, p_{CAB} \in [0,1]$ with $p_{ABC} + p_{ACB} + p_{BAC} + p_{BCA} + p_{CAB} \leq 1$.

Under the assumption that choice trials are independent, the likelihood of obtaining the sample data in Table 1 is maximized at

p_{ABC}	0.343298
p_{ACB}	0.140951
p_{BAC}	0.345591
p_{BCA}	0.000000
p_{CAB}	0.000000
p_{CBA}	0.170160
Log-likelihood	-271.044906

8.2 Model selection

The Akaike information criterion (AIC) offers an estimate of the information loss when a given model is used to represent the data Akaike (1974). The AIC is equal to $2k - 2\ln(L)$ where k is the number of free parameters and L is the maximum value of the likelihood function for the model. It rewards goodness of fit, measured by the likelihood function, and includes a penalty that increases in the number of parameters to discourage overfitting. The table below shows that the Bayesian Probit has the lowest value of AIC. The AIC for the Bayesian probit is calculated with a penalty for $k = 4$ parameters. Relaxing the constraints imposed by equation (3) can only improve the fit, which means the worst possible AIC obtained by a more general version of the Bayesian Probit is bounded above by 546.53. Hence, the constraint imposed by equation (3) do not affect model selection.

Model	k	loglike	AIC
BProbit	4	-265.86	540.53
RUM	5	-271.04	552.09
Logit	2	-275.75	555.49
Probit	4	-274.32	556.65

9 Out-of-sample prediction

We present out-of-sample prediction exercises using the published data of Lea and Ryan (2015). We estimate the model excluding the choice trials for a particular menu, and compare the predicted choice probabilities to the sampled probabilities in that menu. We do this exercise separately for each menu of alternatives.

Table 9 compares the out-of-sample predictions of our model (labeled BP, for Bayesian probit) with the general random utility model, and with two specific instances of random utility (logit and probit). For example, the first column of Table 9 compares the actual data to the prediction results for menu $\{A, B\}$. Each model is estimated excluding the choice trials for $\{A, B\}$. The second column in Table 9 performs the same exercise for $\{A, C\}$. And so on. An asterisk (*) in the table means the model does not have enough empirical bite to predict a single choice probability. The Bayesian probit clearly performs better than the other models in every instance, except in the case of menu $\{A, B\}$, where every model does poorly. Note that excluding the menu $\{A, B\}$ eliminates any evidence of choice reversals and context-effects—in every instance of the restricted dataset, option A is chosen more often than B. While our model is able to accommodate a choice frequency reversal, the version that is more likely to generate the restricted dataset simply gives A more utility than B (see Appendix), and therefore predicts that A would be chosen more often than B from $\{A, B\}$.

Menu	A	B	A	C	B	C	A	B	C	A	B	$\not C$
Data	.37	.63	.84	.16	.69	.31	.55	.28	.17	.61	.39	–
BP	.64	.36	.74	.26	.73	.27	.57	.35	.08	.65	.35	–
Logit	.64	.36	.69	.31	.81	.19	.44	.43	.13	.49	.51	–
Probit	.64	.36	.57	.43	.83	.17	$*$	$*$	$*$	.46	.54	–
RUM	.61	.39	$*$	$*$	$*$	$*$	$*$	$*$	$*$	.42	.58	–

Table 4: Out-of-sample prediction with each model estimated excluding all choice trials for a single menu of alternatives. We compare observed data of each menu to the pseudo out-of-sample predictions from the Bayesian Probit, the RUM family as a whole, and two particular cases of RUM, Logit and Probit. We restrict the Bayesian probit according to equation (3). Estimation details are presented in the Appendix.

10 Rationality, Monty Hall and evolution

Are the observed choices rational? We draw two main lessons from the analysis. First, the data is compatible with the maximization of the complete and transitive ranking $B \succ A \succ C$, under limited sampling. Hence, the choice data is compatible with the narrow definition of 'rational' used in microeconomic decision theory: there exists a stable (context-independent), complete and transitive preference underlying the observed choices.

Second, we can interpret the choice reversals observed in the lab as a direct consequence of the experimental design. The observed choice behavior is compatible with a decision procedure that maximizes the expected value of the chosen alternative when options are drawn independently from the same population. In contrast, the experimental design did not draw options A, B and C randomly and independently from the population. Instead, these options were carefully *designed* to emulate the types of choice reversals observed in the marketing literature. This is in line with a classic interpretation of many results in behavioral economics: decision makers utilize heuristics or rules of thumb that perform well on average outside of the lab (Tversky and Kahneman, 1974). By carefully manipulating the choice environment in the lab, experiments are able to tease out the biases that result from the employment of these heuristics.

Why can't frogs figure out the lab may be different from a typical choice situation? Amphibians have evolved for millions of years. Any decision procedure that mimics Bayesian updating has an evolutionary advantage by getting closer to maximizing the expected value (or "darwinian fitness") of their mating partners. The behavior of frogs can be seen as procedurally rational, in the sense of being optimal for random encounters with potential mates in nature, while at the same time completely failing to respond to the special conditions of the lab environment.

Our explanation for the optimality of the observed choice behavior is analogous to the solution to the classic Monty Hall problem (Selvin, 1975). In the Monty Hall problem a prize is equally likely to be hidden in one three boxes

A, B, C. The contestant initially points to a box. Monty, the game host, has to open one of the two remaining boxes. Since Monty knows where the prize is, Monty always opens an empty box. Suppose for the sake of the example that the contestant initially pointed to box B, and that Monty opened the empty box C. The problem asks if the contestant should switch from the initial pick B to the unopened box A if given the chance. The answer is yes: the strategy of never switching boxes gives 1/3 probability of winning the prize, while the strategy of always switching boxes results in 2/3 probability of winning.

The optimality of switching boxes in the Monty Hall problem is shown by Bayesian updating. Monty's action reveals no information about the initial pick B, but makes it very easy to compare the two remaining boxes A and C. When all options are identical a priori, a Bayesian decision maker optimally chooses among options that are easier to compare.

Many people have trouble understanding the solution to the Monty Hall problem. Why would frogs be able to solve it? To implement the optimal strategy, frogs don't need the ability to explicitly calculate conditional probabilities. Any heuristic that favors options that are easier to compare leads to better mating choices and better darwinian fitness, just like the pure heuristic of always switching doors wins more often in the Monty Hall problem than the pure heuristic of never switching doors. Evolutionary pressure may therefore lead to a hard-wired bias towards choosing among more options that are easier to compare. This mechanism may at least in part help explain the prevalence of the attraction and compromise effects in experiments involving choices by slime molds, bees, birds, frogs, monkeys, and humans.

11 Related Literature

There is a vast literature documenting systematic frequency choice reversals analogous to the ones found in the frog data, among human subjects. For example, Soltani et al. (2012) find the reversal caused by a phantom decoy alternative, where subjects choose money lotteries on a computer screen in the lab. See Ok et al. (2015) for many references that document context dependent behavior in the choice political candidates, medical decision making, investment problems, the evaluation of job candidates and the contingent evaluation of environmental goods.

The specific choice reversals commonly referred to as the attraction and compromise effect have also been found among Rhesus macaques (Parrish et al., 2015), honeybees, grayjays (Shafir et al., 2002), hummingbirds (Bateson et al., 2002), and even a unicellular slime mold (Latty and Beekman, 2011). The slime mold is noteworthy because it has no brain; it doesn't have a single neuron. Our as-if model of optimizing behavior under informational constraints can be used to explain, fit and predict context-dependent choice behavior for human and non-human subjects. Moreover, it suggests a unified explanation for the prevalence of this behavior among many different species.

Our model incorporates several insights from behavioral economics. The idea that rational subjects make optimal inferences from the menu as an explanation for choice paradoxes is proposed in Kamenica (2008). Many psychological explanations of context dependent choice —see, for example, Bordalo et al. (2013) and Tversky and Simonson (1993)— describe a psychological mechanism through which the objectives of the decision maker are influenced by the set of available choices. In our estimation framework, we are able to incorporate some of these psychological insights by making the distribution of posterior beliefs dependent on the context, while at the same reconciling the choice behavior with the existence of a single, stable utility function.

The problem of extending welfare analysis methods to choices that seem to violate rationality postulates has been studied by Green and Hojman (2007) and Bernheim et al. (2009). Our discrete estimation framework can be used to identify stable, context-independent preferences while accommodating some of the previously interpreted as irrational behavioral phenomena. Hence we extend the scope of traditional welfare analysis to discrete choice applications where choices are context-dependent.

12 Concluding Remarks

The remarkably rich dataset from Lea and Ryan (2015) allowed us to transparently show that the Bayesian probit is a useful alternative to the random utility framework for applications in which choices are context-dependent. Our model is able to accommodate both choice reversals found in the experiment while identifying complete and transitive preferences that underlie the data. This in turn makes new kinds of data amenable to standard welfare analysis. We conclude by discussing two possible extensions of our framework.

Taste heterogeneity

We fit a single utility function to the aggregate choices of hundreds of frogs. Hence we explained the data as arising from the maximization of a single preference over the alternatives. In our explanation of the data, every frog would "agree" (if given sufficient information about the options) that option B has the highest darwinian fitness, option A is in the middle, and option C is the least desirable. This explanation lines up with previous studies in biology: there seems to be no evidence of taste heterogeneity in mating among frogs (Lea and Ryan, 2015). Taste heterogeneity alone is not sufficient to explain the present data (as we showed in Section 4) and it is also not necessary (as Table 3 demonstrates).

In applications with human subjects, however, allowing for taste heterogeneity is natural. In applications with repeated choice trials by the same individual subjects (such as in the experimental work of Soltani et al. (2012)) this can be accomplished by fitting the model individually for each subject. This allows different subjects to exhibit different tastes, different levels of risk aversion etc.

In applications where choices are aggregated among different individuals, the utility can be made conditional on a vector of observed subject characteristics, as is common in the applications of random utility framework. Another possibility is to extend our framework to include an additional error term γ_i for each choice option i:

$$X_i = \mu_i + \gamma_i + \varepsilon_i$$

In this case, the true utility of alternative i is $\mu_i + \gamma_i$. To the econometrician, the term μ_i is a deterministic component of utility, while the random term γ_i captures the heterogeneity of tastes among subjects and other components of the value of option i that are unobserved. The term ε_i represents a random perception error by the decision maker, just as before. The decision maker sees the quality of the alternatives in each choice trial as arising from a random draw from the same population. She chooses as if to maximize the expected value of $\mu_i + \gamma_i$ based on observing the noisy measure of value X_i.

Incorporating observable attributes

The simplicity and richness of the frog dataset allowed us to perform the entire analysis "attribute free". Incorporating a vector of observable characteristics for each subject and each choice option is straightforward. The econometrician postulates and tests additional parametric assumptions, writing the choice probabilities of our model conditional on the vector of observable attributes. For example, preferences can be assumed to have a Cobb-Douglas parametric form on the attribute space; the similarity/substitutability parameter can be assumed to be a function of the cosine of the angle between the alternatives, as in Hausman and Wise (1978); the precision parameter can be assumed to be an increasing function of the time that subjects were allowed to contemplate the alternatives before making a choice; and so on. This allows our model to be used for inference and comparative statics on the effect of each observable attribute.

Endogenous precision: choosing when to stop sampling

Finally, we discuss a possible extension of our framework for neuroeconomic applications in which the analyst observes, in addition to choices, the time that each subject took to reach a decision in each choice trial (see for example, Krajbich et al. (2010)). In this case, we may be interested in explaining, fitting and predicting the joint distribution of observed choices and time. A possible extension of our estimation framework that could be useful in such applications is:

(i) Assume that the information precision parameter p is an increasing function of time the decision maker spends contemplating the alternatives;

(ii) Assume the observed time spent contemplating the alternatives before reaching a decision is endogenously chosen by the decision maker while facing a utility cost of contemplation that is increasing in time.

The framework we proposed in this paper corresponds to a particular family of contemplation cost functions, where cost is constant at zero up a level of precision p, and jumpts to plus infinity afterwards. An alternative model with the same number of free parameters would impose a linear fixed cost $c > 0$ per unit of precision, and endogenize the choice of p in each choice trial as the optimal tradeoff between precision and cost. For an example of extending our framework in this direction, the paper by Fudenberg et al. (Working Paper) develops such a model in the special case of binary choice and zero correlations.

A Similarity and Correlation

Cognitive tasks in experimental psychology and psychophysics ask decision makers to choose the largest geometrical figure, the heaviest object, the loudest sound, the darkest shade of gray, and so on. A common finding in these experiments is that decision makers do a better job discriminating among a pair of options i, j when the difference in value is greater. Another common finding is that, keeping the value of the alternatives i, j constant, the ability to discriminate the options improves when the alternatives are more similar. This regularity is known in the psychology literature at least since the experimental work of Tversky and Russo (1969):

> *The similarity between stimuli has long been considered a determinant of the degree of comparability between them. In fact, it has been hypothesized that for a fixed difference between the psychological scale values, the more similar the stimuli, the easier the comparison or the discrimination between them.*

For example, consider the visual task of choosing the triangle with the largest area in Figures 2 and 3. Decision makers typically find it easier to choose, and make less mistakes, in Figure 3. Triangle j is identical in Figures 2 and 3. Triangle i on the left in Figure 2 is very different from triangle i' on the left in Figure 3, but i and i' have exactly the same area. Hence, the difference in area between i and j in Figure 2 is equal to the difference between i' and j in Figure 3. The triangles in Figure 3 are easier to compare because they are more similar.

For a second visual example, suppose we ask subjects to choose which star has more points in Figures 4 and 5 (I am grateful to David K. Levine for suggesting this example.) The star on the left is the same in both Figures. The start on the right has the same number of points in both Figures. Again, subjects usually find it easier to choose and make less mistakes in Figure 5 where the pair of stars is more similar.

In both examples, our model captures the similarity of the alternatives with the correlation parameter. The more similar pair has a higher value of correlation.

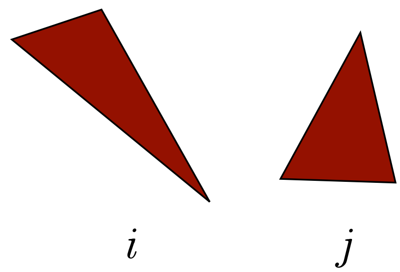

Figure 2: Which triangle has the largest area?

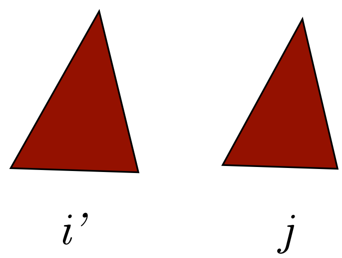

Figure 3: Which triangle has the largest area?

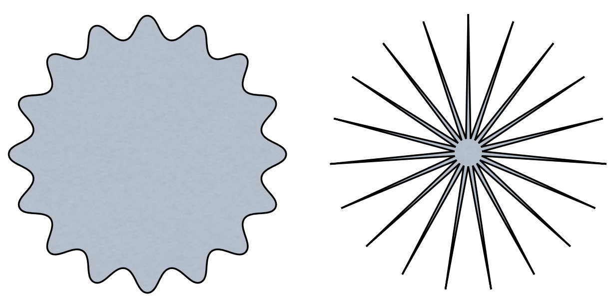

Figure 4: Which star has more points?

Finally, consider an example from a familiar setting of choice problems in economics. Suppose each choice object is a simple lottery (p, m) described by a probability $0 < p < 1$ of winning and a monetary prize $m \geq 0$. Such lotteries are commonly offered in experimental work, e.g. Soltani, De Martino and Camerer

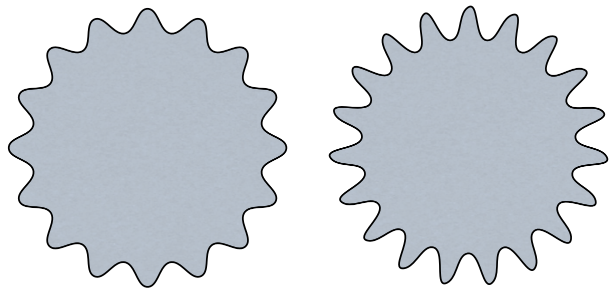

Figure 5: Which star has more points?

(2012). Suppose revealed preference analysis determines for an economic agent that lotteries B, C, D, E are on the same indifference curve while lottery A is superior, as depicted in Figure 6. Tversky and Russo's idea applied to this setting implies the intuitive ranking

$$\rho(A, E) > \rho(A, D) > \rho(A, C) > \rho(A, B) > 1/2.$$

The difference in utility is the same in every pairwise choice, but mistakes are more likely when the options are less similar. Here, similarity means distance according to some metric on an Euclidean space of observed attributes. Fixing the indifference curves, options are harder to compare when they are more distant in the attribute space. In Figure 6, options A and B are the most difficult to compare, while options A and E are the easiest. Note that option A strictly dominates option E, offering a larger prize and a larger probability of winning. The same difference in utility becomes more transparent when options are more similar. Strict dominance can be interpreted as a particularly extreme form of similarity.

B RUM estimates

We use the equivalent RUM formulation of Block and Marschak (1960). Let $p_{ijk} \in [0, 1]$ denote the probability of the strict preference ranking $i \succ j \succ k$

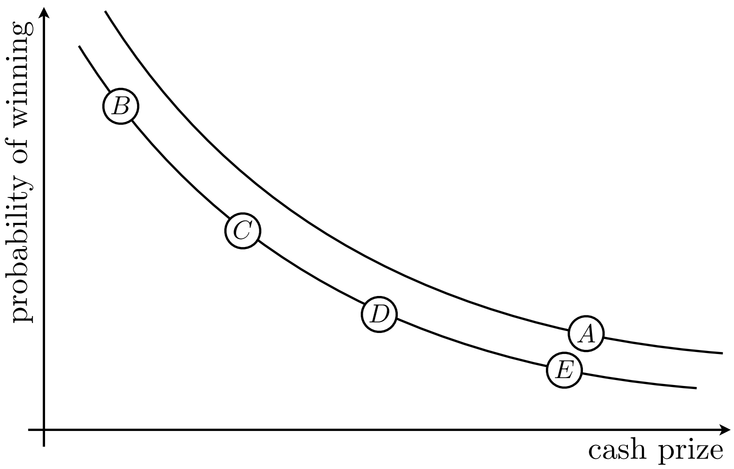

Figure 6: The two indifference curves represent the decision maker's true preferences of over simple money lotteries. Lotteries B, C, D, E lie on the same indifference curve, while lottery A is superior. In pairwise choice tasks, mistakes are more likely when the options are less similar. The comparison of A versus B is the hardest, while the comparison of A and E is the easiest.

for every permutation ijk of A, B and C. Since there is a total of $3! = 6$ such rankings, every RUM can be described by the six non-negative parameters p_{ABC}, p_{ACB}, p_{BAC}, p_{BCA}, p_{CAB}, p_{CBA} with the restriction $p_{ABC} + p_{ACB} + p_{BAC} + p_{BCA} + p_{CAB} + p_{CBA} = 1$. The simplicity of the dataset allows us to obtain solutions to the maximum likelihood problem by solving the standard KKT conditions for an optimum. Results are shown in the table below and we omit standard deviations.

Subsample	Full	$-AB$	$-AB\not{C}$
Estimate p_{ABC}	0.3433	0.4094	0.2880
Estimate p_{ACB}	0.1410	0.1496	0.1356
Estimate p_{BAC}	0.3456	0.2795	0.4009
Estimate p_{BCA}	0.0000	0.0000	0.0000
Estimate p_{CAB}	0.0000	0.0486	0.0000
Estimate p_{CBA}	0.1702	0.1129	0.1756
N Trials	417	299	338
−Loglike	271.04	187.21	214.52

C Logit estimates

We estimate the Logit model in its equivalent formulation by Luce (1959). Choice probabilities are parameterized by three utility values $v_A, v_B, v_C > 0$. The probability that alternative i is chosen from the set Z_1 is given by

$$P(i, Z_1) = \frac{v_i}{\sum_{j \in Z_1} v_j}.$$

With three alternatives, the model has two degrees of freedom. We normalize $v_A = 1$ and estimate v_B and v_C by maximum likelihood. Standard deviations are obtained by inverting the negative of the Hessian matrix at the maximum.

Subsample	Full	$-AB$	$-AC$	$-BC$	$-ABC$	$-AB\not{C}$
Estimate v_B	0.9132	0.5565	1.0148	0.9927	0.9913	1.0554
Std dev	0.1124	0.0950	0.1326	0.1293	0.1310	0.1550
Estimate v_C	0.3045	0.2344	0.4509	0.2304	0.3030	0.3274
Std dev	0.0531	0.0443	0.1039	0.0553	0.0570	0.0584
N Trials	417	299	327	327	377	338
−Loglike	275.75	188.14	233.77	218.12	234.34	221.19

D Probit estimates

The multinomial probit with three alternatives A, B, C is parameterized by a three-dimensional Gaussian distribution with mean and covariance matrix given respectively by

$$\mu = \begin{bmatrix} \mu_A \\ \mu_B \\ \mu_C \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sigma_A^2 & \sigma_A\sigma_B\rho_{AB} & \sigma_A\sigma_C\rho_{AC} \\ \sigma_A\sigma_B\rho_{AB} & \sigma_B^2 & \sigma_B\sigma_C\rho_{BC} \\ \sigma_A\sigma_C\rho_{AC} & \sigma_B\sigma_C\rho_{BC} & \sigma_C^2 \end{bmatrix}$$

There are many redundant parameters in this model. Choice probabilities are unchanged if we subtract the Gaussian variable for option C from the Gaussian variables for A, B, C. Choice probabilities remain unchanged if we further divide all Gaussian variables by the variance of one of them. In other words, the model above is observationally equivalent to a multinomial probit with parameters

$$\tilde{\mu} = \begin{bmatrix} \tilde{\mu}_A \\ \tilde{\mu}_B \\ 0 \end{bmatrix} \quad \text{and} \quad \tilde{\Sigma} = \begin{bmatrix} 1 & \tilde{\sigma}_B\tilde{\rho}_{AB} & 0 \\ \tilde{\sigma}_B\tilde{\rho}_{AB} & \sigma_B^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Hence, only the four parameters $\tilde{\mu}_A, \tilde{\mu}_B, \tilde{\sigma}_B$ and $\tilde{\rho}_{AB}$ can be identified.

Subsample	Full	$-AB$	$-AC$	$-BC$	$-AB\cancel{C}$
Estimate μ_A	0.9397	0.9590	-0.0458	0.9365	0.9457
Std dev	0.1602	0.1541	0.6479	0.1626	0.1612
Estimate μ_B	0.9102	0.6803	0.0009	0.9365	1.1024
Std dev	0.1818	0.2033	0.5897	0.1562	0.2642
Estimate σ_B	1.7265	1.3528	0.0019	0.9999	2.0618
Std dev	0.5394	0.4047	1.1958	0.2792	0.7911
Estimate ρ_{AB}	0.6616	0.8176	0.2893	1.000	0.6064
Std dev	0.3473	0.3180	0.8558	–	0.3532
N trials	417	299	327	327	338
–Loglike	274.32	187.79	231.93	215.25	219.13

E Bayesian probit estimates

We fit the model by maximum likelihood. Recall that we impose the following linear restrictions on the parameter space:

$$\mu_A + \mu_B + \mu_C = 0, \quad \sigma_{AC} + \sigma_{AB} = 1, \text{ and } \sigma_{BC} = 0$$

The restricted version of the model has four degrees of freedom. Estimates using the full sample are presented in the main text. Standard deviations of the estimates are presented for the four estimated parameters. The column header denotes the excluded menu. For example, the column labeled $-AB$ presents the estimates excluding all the choice trials with menu $\{A, B\}$, etc.

Subsample	$-AB$	$-AC$	$-BC$	$-ABC$	$-AB\cancel{C}$
Estimate p	211.3397	0.0533	0.0760	0.1241	0.0751
Std dev	32225.9347	0.0636	0.8087	0.1292	0.0571
Estimate μ_A	0.0308	-0.2678	-0.0147	-0.1956	-0.2135
Std dev	2.3429	0.6083	4.9672	0.4825	0.5661
Estimate μ_B	0.0092	1.6360	1.5918	1.0868	1.3967
Std dev	0.7074	1.1268	11.5579	0.6377	0.6709
μ_C=-$\mu_A - \mu_B$	-0.0401	-1.3683	-1.5771	-0.8912	-1.1832
Estimate σ_{AB}	0.5978	0.0755	0.1234	0.0293	0.0453
Std dev	0.2102	0.1767	2.0767	0.0652	0.0945
$\sigma_{AC} = 1 - \sigma_{AB}$	0.4022	0.9245	0.8766	0.9707	0.9546
$\sigma_{BC} = 0$	0.0000	0.0000	0.0000	0.0000	0.0000
N trials	299	327	327	377	338
–Loglike	187.82	226.23	210.43	225.55	213.23

References

Akaike, Hirotugu, "A new look at the statistical model identification," *Automatic Control, IEEE Transactions on*, 1974, *19* (6), 716–723.

Bateson, Melissa, Susan D Healy, and T Andrew Hurly, "Irrational choices in hummingbird foraging behaviour," *Animal Behaviour*, 2002, *63* (3), 587–596.

Bernheim, B Douglas, Antonio Rangel et al., "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *The Quarterly Journal of Economics*, 2009, *124* (1), 51–104.

Block, H. D. and Jacob Marschak, *Random Orderings and Stochastic Theories of Responses* number 66, Stanford, CA: Stanford University Press,

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Salience and Consumer Choice," *Journal of Political Economy*, 2013.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki, "Stochastic Choice and Optimal Sequential Sampling," Working Paper.

Green, Jerry R. and Daniel A. Hojman, "Choice, Rationality and Welfare Measurement," Working Paper, Harvard University October 2007.

Hausman, Jerry A. and David A. Wise, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 1978, *46* (2), pp. 403–426.

Huber, Joel and Christopher Puto, "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects," *The Journal of Consumer Research*, 1983, *10* (1), pp. 31–44.

_ , J. W. Payne, and C. Puto, "Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis," *Journal of Consumer Research*, 1982, *9* (1), 90–98.

Kamenica, Emir, "Contextual inference in markets: On the informational content of product lines," *The American Economic Review*, 2008, *98* (5), 2127–2149.

Krajbich, Ian, Carrie Armel, and Antonio Rangel, "Visual fixations and the computation and comparison of value in simple choice," *Nature neuroscience*, 2010, *13* (10), 1292–1298.

Latty, Tanya and Madeleine Beekman, "Irrational decision-making in an amoeboid organism: transitivity and context-dependent preferences," *Proceedings of the Royal Society of London B: Biological Sciences*, 2011, *278* (1703), 307–312.

Lea, Amanda M and Michael J Ryan, "Irrationality in mate choice revealed by túngara frogs," *Science*, 2015, *349* (6251), 964–966.

Luce, R. Duncan, *Individual Choice Behavior: a Theoretical Analysis*, Wiley New York, 1959.

Natenzon, Paulo, "Random Choice and Learning," Theoretical Economics, revise and resubmit June 2016.

Ok, E.A., P. Ortoleva, and G. Riella, "Revealed (P)reference Theory," *American Economic Review*, January 2015, *105* (1), 299–321.

Parrish, Audrey E, Theodore A Evans, and Michael J Beran, "Rhesus macaques (Macaca mulatta) exhibit the decoy effect in a perceptual discrimination task," *Attention, Perception, & Psychophysics*, 2015, *77* (5), 1715–1725.

Selvin, Steve, "A Problem in Probability," *The American Statistician*, 1975, *29* (1), 67–71.

Shafir, Sharoni, Tom A Waite, and Brian H Smith, "Context-dependent violations of rational choice in honeybees (Apis mellifera) and gray jays (Perisoreus canadensis)," *Behavioral Ecology and Sociobiology*, 2002, *51* (2), 180–187.

Simonson, Itamar, "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *The Journal of Consumer Research*, 1989, *16* (2), pp. 158–174.

Soltani, Alireza, Benedetto De Martino, and Colin Camerer, "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence," *PLoS computational biology*, 2012, *8* (7), e1002607.

Train, Kenneth, *Discrete Choice Methods with Simulation*, 2nd ed., Cambridge University Press, 2009.

Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, 1974, *185* (4157), 1124–1131.

_ **and Itamar Simonson**, "Context-Dependent Preferences," *Management Science*, 1993, *39* (10), 1179–1189.

_ **and J. Edward Russo**, "Substitutability and similarity in binary choices," *Journal of Mathematical Psychology*, 1969, *6* (1), 1–12.