Random Choice and Learning

Paulo Natenzon

Washington University in Saint Louis

Context-dependent individual choice challenges the principle of utility maximization. I explain context dependence as the optimal response of an imperfectly informed agent to the ease of comparison of the options. I introduce a discrete choice model, the Bayesian probit, which allows the analyst to identify stable preferences from context-dependent choice data. My model accommodates observed behavioral phenomena—including the attraction and compromise effects—that lie beyond the scope of any random utility model. I use data from frog mating choices to illustrate how the model can outperform the random utility framework in goodness of fit and out-of-sample prediction.

I. Introduction

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

$$P(B, \{A, B\}) > .5,$$

This work is based on the first chapter of my doctoral dissertation at Princeton University. It also incorporates and supersedes the working paper "Preference Reversal or Limited Sampling? Maybe Túngara Frogs Are Rational after All." I am grateful to the editor and the anonymous referees for their many useful suggestions that significantly improved the paper. I am indebted to my advisor, Faruk Gul, for his guidance and dedication. I am also indebted to Wolfgang Pesendorfer for many useful discussions. I thank Roland Bénabou, Meir Dan-Cohen, Talia Dan-Cohen, Daniel Gottlieb, David K. Levine, Justinas Pelenis, Carl Sanders, John Yiran Zhu, and audience participants in many seminars for helpful comments and suggestions. All remaining errors are my own. I gratefully acknowledge financial support from the Weidenbaum Center on the Economy, Government, and Public Policy at Washington University in Saint Louis.

Electronically published January 7, 2019
[*Journal of Political Economy*, 2019, vol. 127, no. 1]

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

$$P(A, \{A, B, C\}) > .5.$$

Experimenters have been able to systematically generate such choice reversals in the lab. Famous examples include decoy effects—such as the attraction and the compromise effects—and phantom option effects (Huber, Payne, and Puto 1982; Simonson 1989; Soltani, De Martino, and Camerer 2012). These effects are considered puzzling because they challenge the principle of utility maximization by suggesting that individual preferences are context dependent. Furthermore, they are incompatible with random utility models, which are the models most commonly used in discrete choice estimation.

This paper makes two contributions: first, I provide a new explanation for context dependence in individual choice. In the model, the decision maker is imperfectly informed about the value of the options. Context dependence arises from the optimal response of the decision maker to the ease of comparison of the options, under imperfect information. Incorporating insights from psychology, I model ease of comparison as depending on three factors: the true value of the options, the similarity of the options, and the precision of the information obtained by the decision maker before making a choice. The second main contribution is to provide a new discrete choice estimation framework in which the analyst can identify preferences, similarity, and information precision from choice data. I include an empirical application to illustrate how the new framework can outperform random utility models in goodness of fit and out-of-sample prediction, thus presenting a useful alternative for inference from context-dependent choice data.

The following example shows how ease of comparison affects choice when the decision maker faces uncertainty about the value of the options:

Example 1. A consumer is about to encounter three choice options A, B, and C. The consumer has no information a priori that favors one option over another; her prior beliefs about the utility of the options are given by any exchangeable (i.e., invariant to permutations) and absolutely continuous distribution. Before having a chance to contemplate the three options, she assigns probability 1/6 to each possible strict ranking involving A, B, and C, namely,

$$B \succ A \succ C, \quad B \succ C \succ A,$$
$$A \succ B \succ C, \quad C \succ B \succ A,$$
$$A \succ C \succ B, \quad C \succ A \succ B.$$

Now suppose she obtains some information about the options (e.g., by inspecting the three options) that (i) conveys very little or nothing about the individual values of A, B, and C and (ii) conveys with a high degree of certainty that option A is better than option C. To make it extreme, suppose she learns that the event $(A \succ C)$ occurred and nothing else. Updating the prior with this information will result in probability 1/3 given to each of

$$B \succ A \succ C,$$

$$A \succ B \succ C,$$

$$A \succ C \succ B,$$

and hence option A now has 2/3 probability of being the best option. If she had learned $(C \succ A)$, then C would have been the favored option. Starting from an exchangeable prior, any information that is very imprecise about the individual values of A, B, and C but allows a precise comparison between options A and C will make option B unlikely to be chosen.[1] □

Random utility models can incorporate the ease of comparing options in binary choices, but they cannot account for the intuitive behavior involving the three options presented in example 1. In the random utility framework, the utility of the options is given by random variables U_A, U_B, $U_C : \Omega \to \mathbb{R}$ on a probability space $(\Omega, \mathbb{P})$, and the decision maker chooses the option with the highest random utility realization in each choice trial. In simpler formulations (such as logit) these variables are independent. More flexible formulations (such as nested logit) allow U_A, U_B, and U_C to be correlated. No matter if the variables are correlated or independent, for every state of nature $\omega \in \Omega$ in which $\{U_A > U_B \text{ and } U_A > U_C\}$ we have, in particular, that $\{U_A > U_B\}$. Therefore, the probability of these two events must satisfy

$$\mathbb{P}\{U_A > U_B \text{ and } U_A > U_C\} \leq \mathbb{P}\{U_A > U_B\},$$

and the probability of choosing A can only decrease when option C is introduced. This property, called *monotonicity*, is satisfied by every random utility model, no matter what distributional assumptions we make on U_A, U_B, and U_C. In particular, neither independent shocks nor correlated shocks to utility are able to accommodate the choice reversals I have described.

To explain how the new model departs from random utility, one can interpret the classic random utility framework as an "as-if" model of optimizing behavior subject to informational constraints. In this interpretation, the random variables U_A, U_B, and U_C represent a noisy signal about the true value of each option. The decision maker observes a signal realization for each option in the choice problem and chooses the option with the highest signal realization. This as-if story in which the value of the options is

[1] The optimality of choosing among the pair of options that is easy to compare in this example is analogous to the optimality of switching doors in the solution to the Monty Hall problem (Selvin 1975).

perceived with noise is the interpretation originally proposed for random utility models in psychology (Thurstone 1927a, 1927b).

The new model, the *Bayesian probit*, is parameterized like the classic multinomial probit model: signals have a joint Gaussian distribution. As in the multinomial probit, the signals for different options may be correlated. But this model departs from the random utility framework in two ways. First, the decision maker has a prior. She sees the value of the options in every choice problem as random draws from the same population. Second, the decision maker does not choose the option with the highest signal realization. Instead, she chooses the option with the highest expected value, conditional on the information conveyed by the signals. As a consequence, in this model, the presence of option C offers additional information that can influence the probabilities of choosing A versus B. Example 1 above is an extreme case in which alternative B is never chosen. The Bayesian probit model captures this extreme example when the signals for A and C are perfectly correlated. But the model can also accommodate less extreme cases often found in experimental data.

The experimental literature contains many examples of choice reversals among humans (see Sec. II). In recent years, biologists have also found evidence of choice reversals among a diverse array of species. It turns out, for example, that the same experimental designs that generate choice reversals among human subjects work equally well for monkeys, frogs, birds, bees, and even slime mold. Throughout the paper, I use a data set from Lea and Ryan (2015) involving mate choices by female túngara frogs in the lab for illustrative purposes. Lea and Ryan's data set, which contains two kinds of choice reversals, is simple, yet impressively thorough, insofar as the authors tested all binary comparisons. It is therefore well suited to the task of illustrating the strength of the model with a concrete example. At the same time, the effectiveness of the model in explaining behavior among these nonhumans suggests its broad reach and applicability to contexts that might be of interdisciplinary interest. Indeed, where Lea and Ryan conclude that the frogs are acting irrationally, thereby defying some basic tenets of natural selection, I show that the frogs may have rational preferences after all.

The plan of the paper is as follows. Section II reviews the evidence of context-dependent individual choice in many settings and presents the data set that acts as my exemplary case. Section III provides a new explanation for context dependence as an optimal response of decision makers to imperfect information about the values of the options and to their ease of comparison. In Section IV, I define the primitives of the model and I propose nonparametric definitions of easier to compare, revealed preference, and revealed similarity. Section V introduces the parametric model, the Bayesian probit, explains how the parameters of the model are identified from choice data, and provides testable implications. Section VI contains the main theoretical results, showing how the Bayesian probit accommodates many types of apparently irrational choice reversals from the maximization of a single, stable utility function under informational constraints. In Section VII, I fit the model to the frog mating choice data and perform pseudo out-of-sample prediction exercises. These provide an illustration of how the model can outperform the random utility framework in goodness of fit and prediction. Section VIII concludes the paper by suggesting two possible extensions of the framework.

II. Context-Dependent Choice Data

Experimental work in economics, marketing, and psychology has generated a vast literature documenting context dependence in individual decisions. A clear example is the choice frequency reversal between two options A and B:

$$P(B, \{A, B\}) > .5 \quad \text{and} \quad P(A, \{A, B, C\}) > .5. \tag{1}$$

Option B is chosen in more than half of the choice trials involving binary comparisons between A and B, while option A is chosen in more than half of the choice trials in ternary comparisons among A, B, and C.

Some experimental designs have been shown to be particularly successful at generating choice reversals. The most famous example is perhaps the *attraction effect*. In the attraction effect, the choice reversal in (1) is induced by introducing an asymmetrically dominated decoy option: let options A, B, and C be differentiated in two or more dimensions, let option A dominate the decoy C 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 a two-dimensional setting.

Another famous experimental design that produces the choice reversal in (1) involves an extreme decoy option, such as option C in figure 1. This choice reversal is called the *compromise effect*, because the choice of the middle option A may be justified as a "compromise" between the extreme options B and C.

The decoy alternative has also been shown to induce choice reversals when it is presented for comparison but is not available for choice. In the experiment of Soltani et al. (2012), for instance, subjects compared three options (lotteries over money) on a computer screen for 8 seconds. Next, one of the three options was randomly erased from the choice set, and subjects had 2 seconds to pick one of the remaining alternatives. This *phantom option* design induced significant choice reversals, compared to a separate treatment in which only two options were shown.

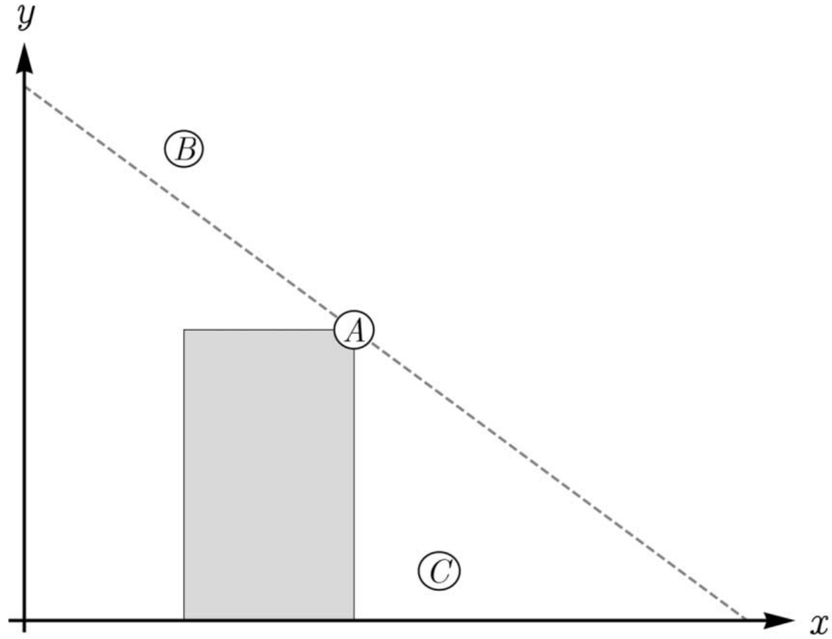

FIG. 1.—Male frog calls *A*, *B*, and *C* are differentiated along two dimensions. The *x*-axis represents a measure of "static attractiveness," while the *y*-axis represents speed (see the supplemental online appendix in Lea and Ryan [2015] for details). The gray rectangle represents the location of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

These experimental designs have been shown to induce choice reversals in a broad range of settings, such as in consumer choices, choice of political candidates, medical decision making, investment problems, job candidate evaluation, and the contingent evaluation of environmental goods.[2] In addition, biologists have shown that the attraction and compromise effects extend to the animal kingdom. Various studies have identified mate or food choice reversals among monkeys, frogs, bees, birds, and even a brainless amoeboid type of mold.[3]

Lea and Ryan (2015) provide an excellent example of choice reversals among female túngara frogs, who choose mating partners on the basis of

[2] The literature on decoy effects originated in marketing (Huber et al. 1982; Simonson 1989). Ok, Ortoleva, and Riella (2015) provide references to evidence in many other areas. See also the discussion of the replicability of these effects in Frederick, Lee, and Baskin (2014), Huber, Payne, and Puto (2014), Simonson (2014), and Yang and Lynn (2014).

[3] Attraction and compromise effect have been found among rhesus macaques (Parrish, Evans, and Beran 2015), túngara frogs (Lea and Ryan 2015), honeybees, gray jays (Shafir, Waite, and Smith 2002), hummingbirds (Bateson, Healy, and Hurly 2002), and a unicellular slime mold (Latty and Beekman 2011).

TABLE 1
MATE CHOICES BY FEMALE TÚNGARA FROGS

Presented Options	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 $\mathbb{C}$	79	.61	.39	. . .

NOTE.—Relative frequencies of choice by female túngara frogs in the data set 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).

the sound of their call. They simulate three different male frog calls in the lab. Following the conventions of the marketing literature, they labeled the options target (A), competitor (B), and decoy (C). Figure 1 shows that the options are differentiated as in the classic compromise effect.[4] Option C is closer to A than to B along every dimension but is not strictly dominated.

The first three rows of table 1 show how often female frogs chose each option 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 reveal a complete and transitive ranking of the three options: $B \succ A \succ C$.

First choice reversal.—The fourth row in table 1 shows that frogs were more likely to choose A over B when all three options 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 can hear the sound of the calls by males A, B, and C when C is a phantom option. This means that, while the three options were equidistant from the female 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 equally well, but only A and B were choosable. Comparing the first and last rows of table 1, we find that the presence of the phantom option C significantly reversed the propensity to choose A over B.

As Lea and Ryan (2015, 965) point out, the data are incompatible with the maximization of utility in the random utility framework: "Female túngara

[4] The sounds 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).

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."

III. Preference, Similarity, and Ease of Comparison

Ease of comparison is crucial to understanding choice reversals once we relax the perfect information assumption of the standard rational choice model. Subjects choose under imperfect information in many settings in which choice reversals are found. In lab experiments with human subjects, the time available to contemplate the options is often manipulated or restricted (see, e.g., Caplin, Dean, and Martin 2011; Soltani et al. 2012). In nature, 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 option before making a choice.

A decision maker that chooses under limited sampling is prone to making choice mistakes. The experimental literature in psychology suggests two important regularities relating the rate of choice mistakes to the value of the options and to the similarity of the options. Cognitive choice tasks in experimental psychology ask participants to choose the largest geometrical figure, the heaviest object, the loudest sound, the darkest shade of gray, and so on among a finite set of available options. A basic finding in these experiments is that decision makers do a better job discriminating among a pair of options i and j when the difference in value is greater.

A second, more subtle finding from experimental psychology is that, for any given fixed difference in the value of options i and j, the ability to discriminate among the options improves when the options are more similar. This regularity has been known in the psychology literature at least since the experimental work of Tversky and Russo (1969, 4): "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."

Figures 2 and 3 illustrate this regularity with two visual task examples. First, consider the visual task of choosing the triangle with the larger area in figures 2*A* and 2*B*. Tversky and Russo (1969) show that in such cases decision makers find it easier to choose, and make fewer mistakes, among the more similar shapes. The triangle on the right is identical in figures 2*A* and 2*B*. The triangle on the left in figure 2*A* is very different from the triangle on the left in figure 2*B*, but both have exactly the same area. Hence,

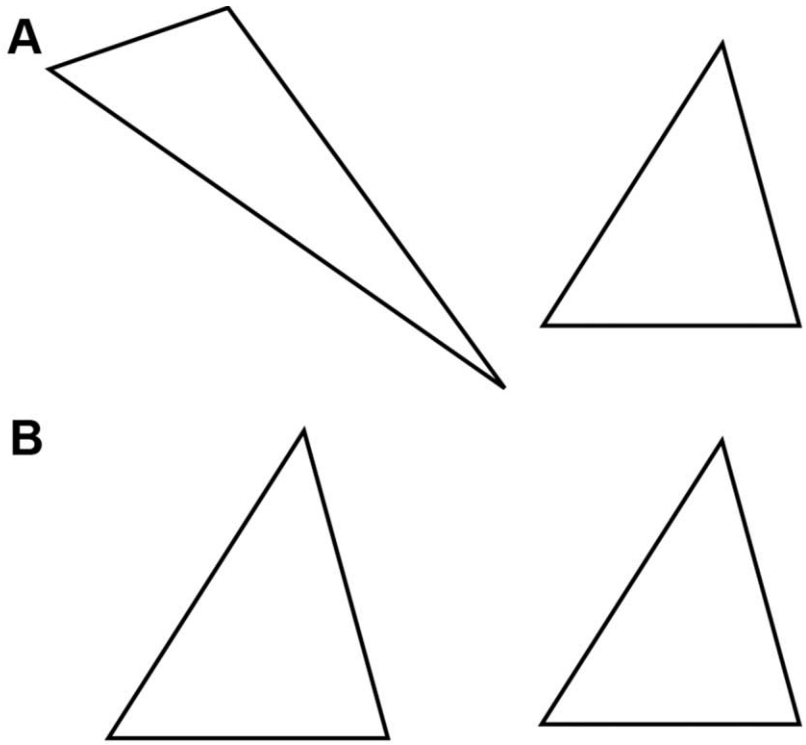

FIG. 2.—Which triangle has the larger area?

the difference in area between left and right in figure 2*A* is equal to the difference in area between left and right in figure 2*B*. The triangles in figure 2*B* are easier to compare because they are more similar.

A different visual task asks subjects which star has more points in figures 3*A* and 3*B*.[5] The star on the left is the same in both figures. The star on the right has the same number of points in both figures. Again, subjects usually find it easier to choose and make fewer mistakes in figure 3*B*, where the pair of stars is more similar. The similarity in figure 3*B* allows a clear relative comparison of the number of points, even without a thorough investigation of the individual values. This can be paraphrased as "I couldn't quite tell the number of points on each star, but I am pretty certain that the star on the right has more points."

These two insights about values and similarity shed light on how the typical experimental designs for the attraction and compromise effects manipulate the ease of comparison of the options. For example, consider

[5] I am grateful to David K. Levine for suggesting this example.

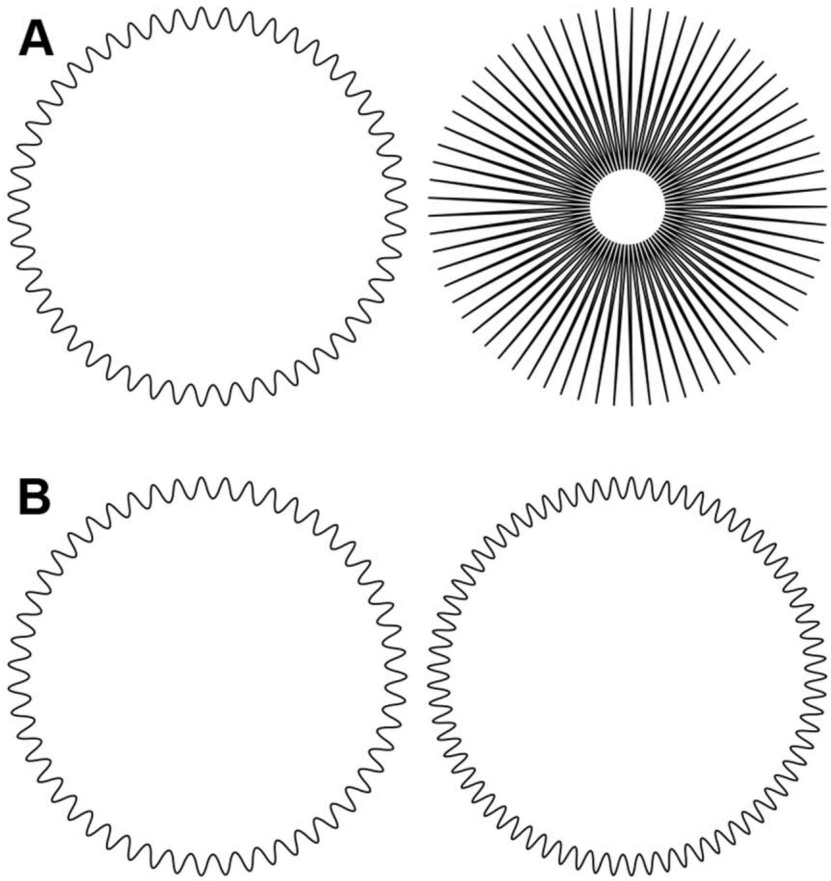

Fig. 3.—Which star has more points?

the three male frog calls used in Lea and Ryan (2015). Figure 1 shows that the decoy *C* is closer to the target *A* than to the competitor *B* in both dimensions. If options *A* and *B* have comparable values, Tversky and Russo's hypothesis, quoted above, says that the decoy *C* should be easier to compare to *A* than to *B*. The revealed preference and revealed similarity analysis that I introduce in the next section formalizes this intuition and shows that in the data the decoy option *C* is revealed more similar to *A* than to *B*.

Example 1 in the introduction shows that with imperfect information and an exchangeable prior, option *B* may be chosen less often in the presence of *C*. If the quality of mating opportunities in nature is close to the quality of random draws from the population, then the choice reversals in the data can be seen as an optimal response to the ease of comparison of the options.

IV. Revealed Preference and Revealed Similarity

In this section, I introduce nonparametric definitions of *easier to compare*, *revealed preference*, and *revealed similarity* and illustrate these nonparametric definitions using the choice frequency data from table 1. To do this, I first define a stochastic choice function P on a domain of choice problems that may include phantom options. The function P is our main modeling primitive: it describes observable choice behavior from the point of view of the analyst.

Let $\mathcal{A} \neq \varnothing$ be the universal set of choice options. For example, in the frog mating choice experiment, $\mathcal{A}$ is the entire population of male túngara frogs. A *choice problem* is a pair (N, M) of disjoint, finite subsets of $\mathcal{A}$, where $N = \{i_1, i_2, \ldots, i_n\}$ is a nonempty set containing $n \geq 1$ choice options and $M = \{j_1, j_2, \ldots, j_m\}$ is a (possibly empty) set containing $m \geq 0$ phantom options. The interpretation of a choice problem (N, M) is that every option in $N \cup M$ is presented to the decision maker for comparison, but only one of the nonphantom options in N must be chosen.

Let $P(i, N, M)$ denote the probability of choosing option i in the choice problem (N, M). The real function P is nonnegative and satisfies

$$\sum_{i \in N} P(i, N, M) = 1$$

in every choice problem (N, M). To simplify notation, I write

$$P(i_k, \{i_1, \ldots, i_n, \not{j}_1, \ldots, \not{j}_m\}) \coloneqq P(i_k, \{i_1, \ldots, i_n\}, \{j_1, \ldots, j_m\}),$$
$$P(i_k, \{i_1, \ldots, i_n\}) \coloneqq P(i_k, \{i_1, \ldots, i_n\}, \varnothing).$$

For example, $P(A, \{A, B, \not{C}\})$ denotes the probability that option A is chosen in the frog mating choice experiment in which the voices of males A, B, and C can be heard equally well in the experimental chamber, but option C is placed on the ceiling and cannot be reached.

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

$$|P(i, \{i, j\}) - .5| > |P(k, \{k, \ell\}) - .5|.$$

A pair is easier to compare whenever the choice frequencies in the binary comparison are more extreme (i.e., closer to zero or one). The first three rows in table 1 show that $\{A, C\}$ is easier to compare than $\{B, C\}$ and $\{B, C\}$ is easier to compare than $\{A, B\}$.

Definition. An option i is *revealed preferred* to j if $P(i, \{i, j\}) > .5$.

Write $i \succ j$ when i is revealed preferred to j, and write $i \succcurlyeq j$ if $j \not\succ i$. Take, for example, binary choice frequencies on the first three rows of table 1. The binary choice frequencies reveal the complete and transitive ranking $B \succ A \succ C$.

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

i. $\{i, j\}$ is easier to compare than $\{k, \ell\}$ and
ii. $k \succcurlyeq i \succ j \succcurlyeq \ell$, with $\succcurlyeq$ transitive.

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

According to the choice frequency data presented in the first three rows of table 1, we have (i) the pair $\{A, C\}$ is easier to compare than any other pair, and, in particular, $\{A, C\}$ is easier to compare than $\{B, C\}$; and (ii) the revealed preference relation is $B \succ A \succ C$. Hence, 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.

In the next section I introduce the parametric model, the Bayesian probit. In this model, the revealed preference relation defined above is captured by a utility function, and the revealed similarity relation is captured by the correlation of perceptual errors. The parametric model can accommodate extreme cases of context-dependent choice behavior illustrated in example 1 of the introduction (proposition 6) but also allows us to fit and predict less extreme cases often found in experimental data (Sec. VII).

V. Parametric Model: The Bayesian Probit

The *Bayesian probit* is a stochastic choice function $P_{\mu,\sigma,p}$ parameterized by a utility function μ, a correlation function σ, and a precision parameter $p > 0$. The true preferences of the decision maker over the universal set of choice options $\mathcal{A}$ are given by a utility function $\mu : \mathcal{A} \to \mathbb{R}$, and μ_i denotes the utility of option i. The decision maker chooses under incomplete information about μ. From the point of view of the decision maker, every choice problem is a collection of options with values μ_i independently drawn from the same Gaussian distribution. With reference to the frog mating choice example, this amounts to assuming that female frogs see every mating 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 some mean m and variance $1/s$. Proposition 1, below, shows that assuming $m = 0$ and $s = 1$ entails no loss of generality.

When presented with a choice problem (N, M), the decision maker observes a noisy signal about utility $X_i = \mu_i + \varepsilon_i$ for each option $i \in N \cup M$. The random vector of signals (X_i) represents how much the decision maker is able to learn about the value of the options in the choice problem before making a choice. The decision maker updates the prior according to Bayes's rule and chooses the option $i \in N$ with the highest posterior mean belief $\mathbb{E}[\mu_i | (X_i)_{i \in N \cup M}]$ among the nonphantom options.

For every finite set of options $\{1, 2, \ldots, n\}$, the signals $(X_1, \ldots, X_n)$ have a joint Gaussian distribution. Each signal $X_i = \mu_i + \varepsilon_i$ corresponding to an option i is equal to true utility μ_i plus a Gaussian perception error ε_i with $\mathbb{E}[\varepsilon_i] = 0$. All perception errors have the same variance $1/p$ given by a single precision parameter $p > 0$. I allow perception errors to be correlated and let $\sigma_{ij} := \mathrm{Corr}(\varepsilon_i, \varepsilon_j)$ for all i, j. The correlation parameters satisfy

$\sigma_{ij} = \sigma_{ji}$ and $\sigma_{ii} = 1$ for every i, j. In addition, the determinant of the correlation matrix $(\sigma_{ij})_{i,j=1,\ldots,n}$ is strictly positive for every finite subset of options $\{1, \ldots, n\}$, which implies that the joint Gaussian distribution of signals is nondegenerate in every choice problem.

In summary, in the Bayesian probit $P_{\mu,\sigma,p}$ with utility μ, correlation σ, and precision p, the probability of choosing each $i \in N$ in choice problem (N, M) is given by

$$P_{\mu,\sigma,p}(i, N, M) \coloneqq \mathbb{P}\left\{ \mathbb{E}\left[\mu_i - \mu_j | (X_k)_{k \in N \cup M}\right] \geq 0 \text{ for all } j \in N \right\}. \quad (2)$$

A. *Identification of Parameters*

Next, I describe the identification properties of the Bayesian probit model. Proposition 1 shows that the prior may be assumed to be standard Gaussian without loss of generality. Proposition 2 identifies ordinal information about μ and σ from binary choice data. Finally, I explain how data from choice problems involving three options suffice to locally identify all the parameters.

PROPOSITION 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 $\sigma_{ij} = \hat{\sigma}_{ij}$.

Normalizing the prior to be standard Gaussian, proposition 1 shows that utility and precision parameters have a cardinal interpretation. The utility parameters $\mu_i = \sqrt{\hat{s}}(\hat{\mu}_i - \hat{m})$ are measured in standard deviations from the population mean. For example, an estimate of $\mu_i = -0.5$ means that the Darwinian fitness of male frog i is half a standard deviation below the population mean. The precision parameter $p = \hat{p}/\hat{s}$ measures the precision of the information obtained by the decision maker in units of precision in the population distribution. This is different from the random utility framework, where utility parameters have only ordinal (but not cardinal) meaning.

PROPOSITION 2. Let the stochastic choice function $P = P_{\mu,\sigma,p}$ be a Bayesian probit with utility μ, correlation σ, and precision $p > 0$. Then

i. option i is revealed preferred to j if and only if $\mu_i > \mu_j$;
ii. the pair $\{i, j\}$ is revealed more similar than $\{k, \ell\}$ only if $\sigma_{ij} > \sigma_{k\ell}$.

Proposition 2 uses binary choice data to identify ordinal information about μ and σ. Part i shows that revealed preference is represented by the utility function μ_i. Hence, the weak revealed preference relation $\succcurlyeq$ obtained from the Bayesian probit is complete, transitive, and independent of correlation and precision parameters. Part ii shows that the revealed similarity relation is captured by the correlation parameters σ_{ij}. Since the revealed similarity relation may not be complete, we need more than binary choice data to identify the full similarity ranking.

In the online appendix, I formally show that the mapping

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{23} \\ p \end{bmatrix} \mapsto \begin{bmatrix} P_{\mu,\sigma,p}(1,\{1,2\},\varnothing) \\ P_{\mu,\sigma,p}(1,\{1,3\},\varnothing) \\ P_{\mu,\sigma,p}(2,\{2,3\},\varnothing) \\ P_{\mu,\sigma,p}(1,\{1,2\},\{3\}) \\ P_{\mu,\sigma,p}(1,\{1,3\},\{2\}) \\ P_{\mu,\sigma,p}(2,\{2,3\},\{1\}) \\ P_{\mu,\sigma,p}(1,\{1,2,3\},\varnothing) \\ P_{\mu,\sigma,p}(2,\{1,2,3\},\varnothing) \end{bmatrix}$$

is locally invertible at almost all points in the parameter space. Hence, data from every choice problem involving three options are sufficient to locally identify all the parameters. The identification is *generic*: it holds for parameter values that lie outside a set of Lebesgue measure zero in the parameter space.

In applications, data are often limited. In such cases, the econometrician achieves identification by formulating and testing additional restrictions on the space of parameters. In experimental work involving choice over lotteries, for instance, analysts often employ the assumption that subjects have expected utility preferences with constant absolute risk aversion. Under this assumption, a single risk aversion parameter determines the utility of every option. This greatly reduces the dimensionality of the parameter space and obviates the need to estimate separate utility values μ_i for every option i. In Section VII, I illustrate how point identification is achieved through parameter restrictions using the experimental frog mating choice data.

B. Testable Implications

The Bayesian probit model carries a set of restrictions for stochastic choice behavior.[6] These nonparametric conditions are directly testable on choice data.

[6] The Bayesian probit lies outside the scope of recent axiomatic work on stochastic choice. Gul, Natenzon, and Pesendorfer (2014), Manzini and Mariotti (2014), and Lu (2016) characterize behavior nested in the random utility framework. The main condition satisfied by Fudenberg, Iijima, and Strzalecki (2015) is ordinal independence from irrelevant alternatives, an analogue of the IIA property of the logit model. In particular, none of these models can accommodate the kinds of context-dependent behavior that I use the Bayesian probit to address in the next section.

DEFINITION. A stochastic choice function P satisfies *moderate transitivity* if

$$\left.\begin{array}{l} P(i,\{i,j\}) \geq .5 \\ P(j,\{j,k\}) \geq .5 \end{array}\right\} \Rightarrow P(i,\{i,k\}) \geq \min\{P(i,\{i,j\}), P(j,\{j,k\})\}. \quad \text{(MT)}$$

Suppose option i is chosen more often than j and option j is chosen more often than k in binary comparisons. Moderate transitivity requires that i is chosen more often than k in a binary comparison, and, moreover, the decision maker cannot do a worse job discriminating the options in $\{i, k\}$ than in both $\{i, j\}$ and $\{j, k\}$.

DEFINITION. The function P satisfies the *Block-Marschak inequalities with phantom options* if for every finite set of options N' and every $i \in N \subseteq N'$,

$$\sum_{M : N \subseteq M \subseteq N'} (-1)^{|M \setminus N|} P(i, M, N' \setminus M) \geq 0. \quad \text{(BM-P)}$$

DEFINITION. The function P satisfies *moderate transitivity with phantom options* if, for any three distinct options i, j, k and any finite set of options M distinct from i, j, k,

$$[P(i,\{i,j\},\{k\} \cup M) \geq .5 \text{ and } P(j,\{j,k\},\{i\} \cup M) \geq .5]$$

$$\text{implies} \quad \text{(MT-P)}$$

$$P(i,\{i,k\},\{j\} \cup M) \geq \min\{P(i,\{i,j\},\{k\} \cup M), P(j,\{j,k\},\{i\} \cup M)\}.$$

PROPOSITION 3. The function P is a Bayesian probit only if it satisfies MT, MT-P, and BM-P.

The proof of proposition 3, in the appendix, shows that conditions MT and MT-P are consequences of the assumption that the prior and the signals have a Gaussian distribution, while condition BM-P must also hold in a generalized model that assumes any other prior and signal distributions.

Conditions MT, MT-P, and BM-P are directly testable on choice data. The frog mating choice experiment includes all pairwise comparisons among options A, B, and C and provides a test of moderate transitivity (MT). Table 1 shows that the data support this prediction:

$$P(B,\{B,C\}) = .69 > .63 = \min\{P(B,\{A,B\}), P(A,\{A,C\})\}.$$

Condition BM-P implies a monotonicity property with respect to phantom options: when a choice option becomes a phantom option, the choice probability of the remaining choice options cannot decrease. The frog mating choice experiment allows us to test two of these inequalities:

$$P(A,\{A,B,C\}) \leq P(A,\{A,B,\not{C}\}),$$
$$P(B,\{A,B,C\}) \leq P(B,\{A,B,\not{C}\}).$$

Both predictions are supported by the data: Table 1 shows that the probability of choosing A increases from .55 to .61 and the probability of choosing B increases from .28 to .39 when the choice option C becomes the phantom option $\not{C}$.

While condition MT-P cannot be directly tested on the frog mating choice data, it offers three out-of-sample predictions involving choice problems $\{A, \not{B}, C\}$ and $\{\not{A}, B, C\}$ (using the value $P(A, \{A, B, \not{C}\}) = .61$ from table 1):

i. $\min\{P(A, \{A, \not{B}, C\}), P(C, \{\not{A}, B, C\})\} \leq .61$;
ii. if $P(B, \{\not{A}, B, C\}) \geq .5$, then $P(A, \{A, \not{B}, C\}) \geq \min\{.61, P(B, \{\not{A}, B, C\})\}$;
iii. if $P(C, \{A, \not{B}, C\}) \geq .5$, then $P(C, \{\not{A}, B, C\}) \geq \min\{.61, P(C, \{A, \not{B}, C\})\}$.

C. Departure from Random Utility Models

Recall that in the random utility framework, each choice option i is associated with a random variable X_i. An option is chosen when it maximizes random utility:

$$P_{\mathrm{RU}}(i, \{1, \dots, n\}) \coloneqq \mathbb{P}\{X_i \geq X_j \text{ for all } j = 1, \dots, n\}. \tag{3}$$

Random utility models differ only on the assumptions made about the distribution of the X_i's. For example, the logit model assumes independent Gumbel distributed variables, the nested logit model assumes correlated Gumbel distributed variables, the multinomial probit assumes Gaussian variables, and so on.

Among the random utility models, the multinomial probit offers the most relevant comparison to the Bayesian probit model. In the multinomial probit, the X_i's are assumed Gaussian with an arbitrary covariance matrix. Hence, the distribution of signals in the Bayesian probit is a particular case of the distribution of random utilities in the multinomial probit. This comparison highlights the Bayesian updating step as the key departure of the model from random utility maximization.

Example 2 (Naive probit). Suppose that a decision maker observes signals $X_1, \dots, X_n$ with a joint Gaussian distribution parameterized by μ, σ, p as in the Bayesian probit but chooses according to (3) instead of (2). This decision maker ignores the prior and skips the Bayesian updating step. Instead of solving a sophisticated noise filtering problem, this decision maker simply chooses the object that "looks best."[7] Call the result-

[7] This interpretation of naive behavior comes from the study by Weibull, Mattson, and Voorneveld (2007), who study the optimality of naive choice rules in a setting in which perceptual errors are uncorrelated.

ing choice function the *naive probit* with parameters μ, σ, and p and denote it by $\ddot{P}_{\mu,\sigma,p}$. Since it maximizes Gaussian random utility, by definition, the naive probit is a special case of the classic multinomial probit model, assuming the variance is equal for all options. □

The naive probit model shows that one way to describe the departure of the Bayesian probit model from the random utility framework is the substitution of the "sophisticated" choice rule (2) for the "naive" choice rule (3). Likewise, by changing the distributional assumptions on X_i, every random utility model can be interpreted as the choice function of a naive decision maker who observes random signals X_i, ignores the prior, skips the Bayesian updating step, and simply chooses the option with the highest signal realization.

Under certain conditions, the Bayesian probit and the naive probit coincide:

LEMMA 1. If $P_{\mu,\sigma,p}$ is a Bayesian probit and $\ddot{P}_{\mu,\sigma,p}$ is a naive probit, then

i. $P_{\mu,\sigma,p}(i,\{i,j\}) = \ddot{P}_{\mu,\sigma,p}(i,\{i,j\})$ for all i, j, μ, σ, and p;

ii. $P_{\mu,\sigma,p} = \ddot{P}_{\mu,\sigma,p}$ if σ is constant;

iii. $\lim_{p\to\infty}[P_{\mu,\sigma,p}(i,N,M) - \ddot{P}_{\mu,\sigma,p}(i,N,M)] = 0$ for all i, N, M, μ, and σ.

In parts i and ii, the complete symmetry of perceptual errors makes the signals equally informative about every option: an option has the highest posterior mean belief if and only if it has the highest signal realization. This hinges on the assumption that every signal has the same precision.[8] Part iii shows that the Bayesian and the naive probit exhibit the same asymptotic choice behavior with arbitrarily high information precision. It is easy to see that with arbitrarily precise signals the decision maker chooses the best alternative with probability 1 in both models. But part iii says more: in the knife-edge case in which more than one choice option ties for maximum utility, both models break the tie in exactly the same way.

Part i implies the binary choice formula

$$P_{\mu,\sigma,p}(i,\{i,j\}) = \Phi\left(\frac{\sqrt{p}}{\sqrt{2}} \times (\mu_i - \mu_j) \times \frac{1}{\sqrt{1-\sigma_{ij}}}\right), \tag{4}$$

where Φ denotes the cumulative distribution function of the standard Gaussian distribution. Equation (4) shows that the ability of the decision maker to compare and correctly discriminate among the pair of options i, j improves with increases in the precision of the signals p, in the difference in value $|\mu_i - \mu_j|$, and in the signal correlation σ_{ij}. Increasing the precision parameter p improves discrimination in every binary choice problem. The

[8] If the signal for option i were less precise than the signal for option j, the Bayesian decision maker would give the signal about j more weight, breaking the equivalence. For the same reason, the equivalence breaks down if the decision maker must separately learn the precision of each signal (see the learning model in Harbaugh, Maxwell, and Shue [2016] for an example).

correlation parameter σ_{ij} affects the ability to discriminate between a specific pair of options i, j and can vary independently from their value.

The Bayesian updating step introduces context dependence in the model: the distribution of posterior beliefs about the value of a given option i can be affected by the presence of new options through signal correlation. For example,

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

whenever X_C is correlated with X_A.[9] Hence, the presence of a signal about C offers additional information that can influence the probability of choosing A over B in the model.[10] As I show in the next section, this allows the Bayesian probit to accommodate context-dependent choice behavior that lies outside the scope of the random utility framework.

VI. Accommodating Choice Reversals

In the introduction, example 1 shows that ease of comparison can be an important determinant of choice behavior under imperfect information. In the Bayesian probit model, ease of comparison is determined by the correlation of perceptual errors when information precision is low. For any two options A and C, the difference in signals $X_A - X_C$ is given by

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

where $\mu_A - \mu_C$ is the difference in value between A and C and $\varepsilon_A - \varepsilon_C$ is the difference in perceptual errors. For any fixed signal precision p,

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

is strictly decreasing on the correlation σ_{AC}. In the limit, as correlation σ_{AC} goes to one, $\text{Var}[\varepsilon_A - \varepsilon_C]$ goes to zero. Highly correlated perceptual errors tend to "cancel out," revealing the true difference $\mu_A - \mu_C$ with high precision. If precision p is low and correlation σ_{AC} is high, the signals X_A and X_C reveal very little about the individual values μ_A and μ_C but a great deal about the difference in values $\mu_A - \mu_C$. 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."

[9] Other forms of contextual inference by the decision maker may lead to the same inequality. For example, Kamenica (2008) shows that it is optimal for a consumer to exhibit context-dependent choice behavior if an agent who has private information about the values of the options (e.g., a monopolist firm) has designed the menu of options available in the choice problem.

[10] Many existing explanations of context-dependent choice—see, e.g., Tversky and Simonson (1993), de Clippel and Eliaz (2012), and Bordalo, Gennaioli, and Shleifer (2013)—describe a psychological mechanism through which the objectives of the decision maker are influenced by the set of available choices. The context dependence of posterior beliefs in this model incorporates some of these psychological insights, while at the same time reconciling the choice behavior with the existence of a single, stable utility function.

The rest of this section shows how the interaction between information precision and similarity allows the model to explain several puzzling context effects in choice.

A. *Interaction of Similarity and Information Precision*

When information precision p is large, choice mistakes are unlikely. In that case, similarity has a negligible effect on choice probabilities, unless true utility values are very close (see, e.g., eq. [4]). Proposition 4 shows the effect of the similarity parameter on choice probabilities when it is most relevant, that is, when it acts as a tiebreaker among three equally desirable options.

PROPOSITION 4 (High precision). If $\mu_1 = \mu_2 = \mu_3$ and $\sigma_{23} > \sigma_{13}$, then, for all $p > 0$ sufficiently large,

$$\frac{P_{\mu,\sigma,p}(1, \{1, 2, 3\})}{P_{\mu,\sigma,p}(2, \{1, 2, 3\})} > \frac{P_{\mu,\sigma,p}(1, \{1, 2\})}{P_{\mu,\sigma,p}(2, \{1, 2\})}.$$

In this case, introducing the new option 3 hurts the probability of choosing the similar option 2 proportionally more than 1. Hence, when the decision maker has sufficiently precise information, the similarity parameter breaks ties in the direction predicted by Tversky's (1972, 283) well-known *similarity hypothesis*: "The addition of an alternative to an offered set 'hurts' alternatives that are similar to the added alternative more than those that are dissimilar to it." This so called *similarity effect* can be accommodated by random utility models by making the random value of options correlated (Train 2009). Unlike the random utility framework, however, the Bayesian probit allows the inequality of proposition 4 to be reversed when the decision maker is relatively uninformed:

PROPOSITION 5 (Low precision). If $\sigma_{12} > -1/2$ and $\sigma_{23} > \sigma_{13}$, then, for all $p > 0$ sufficiently small,

$$\frac{P_{\mu,\sigma,p}(1, \{1, 2, 3\})}{P_{\mu,\sigma,p}(2, \{1, 2, 3\})} < \frac{P_{\mu,\sigma,p}(1, \{1, 2\})}{P_{\mu,\sigma,p}(2, \{1, 2\})}.$$

This is a violation of Tversky's similarity hypothesis. It requires two initial options 1 and 2 that are not too dissimilar ($\sigma_{12} > -1/2$) and an added option 3 that is more similar to option 2 than to option 1 ($\sigma_{23} > \sigma_{13}$). For low values of information precision p, the introduction of option 3 hurts the dissimilar option 1 proportionally more than the similar option 2. The statement of proposition 5 holds independently of the true value of the options.

B. *Decoy Effects*

PROPOSITION 6 (Decoy effect). Fix any two options 1 and 2 with similarity parameter σ_{12} and let $\varepsilon > 0$. If an added option 3 is such that

$\sigma_{23} > \sigma_{12}$, $\sigma_{23} > \sigma_{13}$, σ_{23} is sufficiently close to its upper bound, and information precision $p > 0$ is sufficiently small, then $P_{\mu,\sigma,p}(1, \{1, 2, 3\}) < \varepsilon$.

In consumer choice experiments in marketing, options 1, 2, and 3 are called, respectively, the *competitor*, the *target*, and the *decoy*. The inequalities $\sigma_{23} > \sigma_{12}$ and $\sigma_{23} > \sigma_{13}$ can be paraphrased, respectively, as “the target is more similar to the decoy than to the competitor” and “the decoy is more similar to the target than to the competitor.” Proposition 6 shows that, if the similarity between the decoy and target is sufficiently high and information precision is sufficiently low, the competitor is chosen with probability close to zero when the decoy is introduced.

Proposition 7 (Violation of monotonicity). Suppose $P_{\mu,\sigma,p}(1, \{1, 2\}) = 0.5$ and fix any $\varepsilon > 0$. If option 3 is revealed more similar to 2 than to 1 and is sufficiently inferior, then $P_{\mu,\sigma,p}(2, \{1, 2, 3\}) > 0.5$ for all precision values $p > \varepsilon$.

Here, option 1 (the competitor) and option 2 (the target) have the same utility, while the decoy option 3 is inferior. Adding the decoy option 3 increases the probability of choosing the target option 2 above .5. This is a violation of monotonicity, which cannot be accommodated by any random utility model. In particular, proposition 7 shows that it is possible to obtain a violation of monotonicity for any fixed information precision value $p > 0$.

Proposition 8 (Avoidance of extremes). If $\sigma_{12} = \sigma_{23} > \sigma_{13}$, then option 2 is the option chosen most often from {1, 2, 3} for all sufficiently small $p > 0$.

If we interpret the similarity parameter σ_{ij} as a measure of “closeness” between options i and j, then $\sigma_{12} = \sigma_{23} > \sigma_{13}$ means that option 2 is equally “close” to each of options 1 and 3, while options 1 and 3 are less “close” to each other. Proposition 8 says that a sufficiently uninformed decision maker will choose each of the “extreme” options 1 and 3 less often than option 2.

C. *Attraction and Compromise Effects*

The attraction and compromise effects are perhaps the two most well-known decoy effects in the literature (Huber et al. 1982; Simonson 1989). Both effects take place in a setting in which choice options are differentiated along two or more measurable attributes. To show how the Bayesian probit accommodates these classic decoy effects, I link the abstract utility and similarity parameters to the attribute space. For simplicity, I provide an example in which options are differentiated along two dimensions (fig. 4). The extension to attribute spaces with more than two dimensions is straightforward.

I begin with utility. Assume that utility is strictly increasing and strictly concave on the two-dimensional attribute space $\mathbb{R}^2_+$ depicted in figure 4. Options 1 and 2 are located on the same indifference curve, with utility

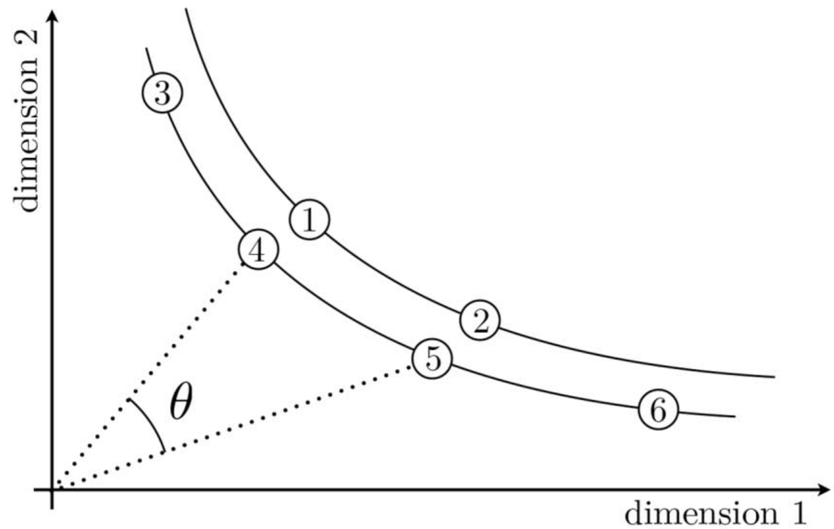

FIG. 4.—Cosine similarity (Hausman and Wise 1978) measures the "closeness" between two options *j* and *k* by the cosine of the angle formed by *j* and *k* in the space of observable attributes. Pairs of options that form wider angles have less correlated perceptual errors.

values $\mu_1 = \mu_2 = 1$. The remaining options lie on a lower indifference curve with utility $\mu_3 = \mu_4 = \mu_5 = \mu_6 = -1$.

Next, I link the abstract similarity parameters to the two-dimensional attribute space. Following the discrete choice estimation literature, I assume that the similarity parameter σ_{ij} is strictly increasing on the "closeness" of options *i* and *j* in the space of measurable attributes (Domencich and McFadden 1975; Hausman and Wise 1978). The exact meaning of "close" depends on the notion used by the econometrician. For example, Hausman and Wise measure "closeness" using *cosine similarity*: the cosine of the angle formed by options *i* and *j* in the space of measurable attributes. In figure 4, the angle formed by options 4 and 5 is denoted by θ. The pairs of options {1, 4} and {2, 5} are the most similar, forming angles with cosines close to one, while the pair {3, 6} is the least similar, forming an angle with the cosine close to zero. For the sake of illustration, I fix similarity values $\sigma_{14} = \sigma_{25} = 0.95$, $\sigma_{12} = \sigma_{15} = \sigma_{24} = 0.7$, and $\sigma_{16} = \sigma_{23} = 0.2$.

Using these numerical values for utility and similarity, the choice options depicted in figure 4 illustrate how the Bayesian probit generates the attraction and compromise effects.

Attraction effect.—Recall that the attraction effect describes a violation of monotonicity that results from the addition of an *asymmetrically dominated* decoy option. An option is asymmetrically dominated when it is strictly dominated by an existing option and undominated by another. Huber et al. (1982) discovered that introducing an asymmetrically dominated decoy option can increase the probability that the dominant option is chosen.

Suppose the initial options are 1 and 2 depicted in figure 4. The decoy option 4 is dominated by option 1 but not by 2. Conversely, the decoy option 5 is dominated by option 2 but not by 1. The Bayesian probit accommodates the attraction effect in both directions. Proposition 5 shows that in either case introducing the decoy hurts the competitor proportionally more than the target for low values of information precision. Proposition 6 shows conditions under which the competitor is chosen with probability close to 0, again for sufficiently low information precision. Proposition 7 shows that monotonicity can be violated for any fixed precision value $p > 0$ if the decoy option has a sufficiently low utility. Figure 5 illustrates the attraction effect with our numerical example. The figure plots the choice probabilities for target, competitor, and decoy as a function of information precision. The figure shows that the violation of monotonicity $P_{\mu,\sigma,p}(1, \{1, 2, 4\}) > P_{\mu,\sigma,p}(1, \{1, 2\})$ (respectively, $P_{\mu,\sigma,p}(2, \{1, 2, 5\}) > P_{\mu,\sigma,p}(2, \{1, 2\})$) holds for a wide range of precision values.

Compromise effect.—Simonson (1989) shows the first evidence for the tendency of a decision maker to avoid extreme options. For example, the introduction of the extreme option 6 in figure 4 favors a choice of option 2 over option 1, while the introduction of option 3 favors the choice of option 1 over option 2. As noted earlier, the term compromise effect

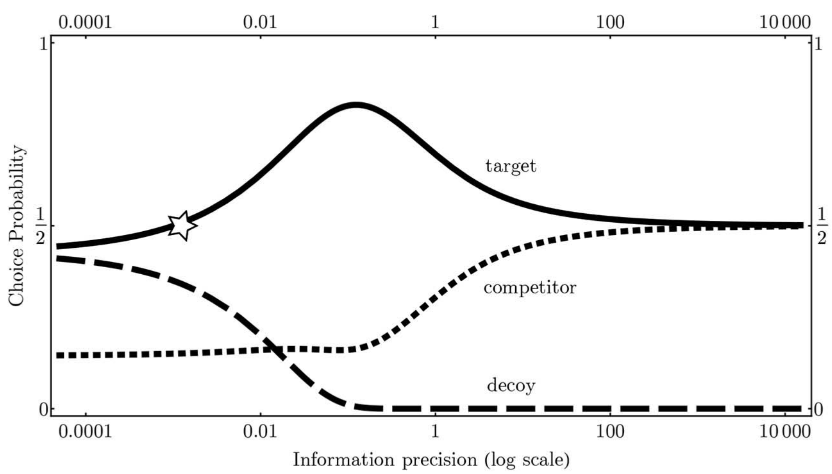

FIG. 5.—The attraction effect. Choice probabilities for target (solid line), competitor (dotted line), and decoy (dashed line) plotted as a function of information precision in logarithmic scale. The star marks the point where the choice probability for the target increases above 1/2, violating monotonicity. Utility parameters are 1.0 for the target and competitor and −1.0 for the decoy. Similarity parameters are 0.7 between the target and competitor, 0.7 between the competitor and decoy, and 0.95 between the target and decoy.

suggests that the middle option is favored because it may be seen as a compromise between the two extreme options. Proposition 8 predicts these effects for small information precision values, and proposition 7 shows that they may lead to violations of monotonicity for any fixed information precision value $p > 0$ if the extreme decoy options have sufficiently low utility. Figure 6 illustrates the compromise effect with the numerical example. Two violations of monotonicity $P_{\mu,\sigma,p}(1, \{1, 2, 3\}) > P_{\mu,\sigma,p}(1, \{1, 2\})$ and $P_{\mu,\sigma,p}(2, \{1, 2, 6\}) > P_{\mu,\sigma,p}(2, \{1, 2\})$ hold for a wide range of precision values.

The attraction and compromise effects illustrated above are robust to small changes to the parameter values. In particular, options 1 and 2 do not need to be indifferent in order to obtain the violations of monotonicity illustrated above. If utility and similarity are assumed to change continuously as a function of the location of the options in the space of attributes, then these effects are also robust to small perturbations to the locations of the options.

Some deterministic models of context-dependent choice can also reconcile the attraction effect with the assumption that preferences are represented by a stable utility function. For example, the decision maker may maximize utility over a subset of options from the original choice problem, called a *consideration set*. If the consideration set excludes the competitor option when the decoy is introduced, a choice reversal may occur. In the imperfect attention model of Masatlioglu, Nakajima, and Ozbay (2012), the consideration set is determined by an attention filter. In the

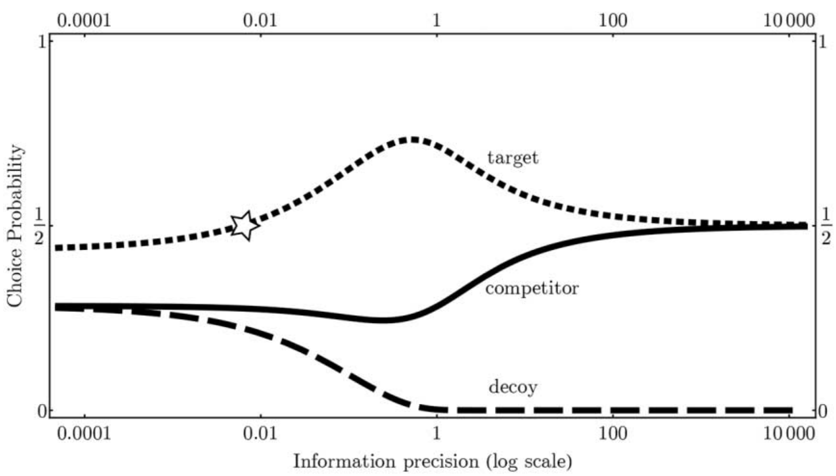

FIG. 6.—The compromise effect. Choice probabilities for target (dotted line), competitor (solid line), and decoy (dashed line) plotted as a function of information precision in logarithmic scale. The star marks the point where the choice probability for the target increases above 1/2, violating monotonicity. Utility parameters are 1.0 for the target and competitor and −1.0 for the decoy. Similarity parameters are 0.7 between the target and competitor, 0.7 between the target and decoy, and 0.2 between the competitor and decoy.

reference dependent model of Ok et al. (2015), the consideration set is determined by a reference option in each choice problem. Both models are deterministic, and the underlying mechanisms of reference dependence and attention filtering are modeled as all-or-nothing phenomena. In contrast, in my stochastic choice model the decision maker pays equal attention to all the options, and the notions of preference, similarity, and information precision are gradual and continuous. One advantage of this approach is that it can provide comparative statics on the intensity of choice reversals based on changes to the attributes of each option. Another advantage is that the framework is directly applicable to discrete choice estimation, as I show next.

VII. Empirical Application

I estimate the parameters of the Bayesian probit model using the frog mating choice data of Lea and Ryan (2015) shown in table 1. The model accommodates all the qualitative features of the data, including the observed choice frequency reversals (table 3). It also outperforms any random utility model with respect to goodness of fit (table 4) and out-of-sample prediction (table 5).

A. Bayesian Probit Estimates

The Bayesian probit model over three options A, B, and C has seven parameters: three utility parameters μ_A, μ_B, and μ_C; three similarity parameters σ_{AB}, σ_{AC}, and σ_{BC}; and one precision parameter p. To obtain identification, I impose the following additional restrictions on the parameter space:

$$\mu_A + \mu_B + \mu_C = 0, \tag{5}$$

$$\sigma_{AC} + \sigma_{AB} + \sigma_{BC} = 1, \tag{6}$$

$$\sigma_{AB} \times \sigma_{AC} \times \sigma_{BC} = 0. \tag{7}$$

Restriction (5) eliminates one utility parameter by assuming that the average utility among the three options used in the experiment matches the average utility in the population. Restrictions (6) and (7) together eliminate two similarity parameters. Equation (6) restricts the average similarity parameter to be 1/3. Equation (7) forces one of the similarity parameters to be 0. The restrictions are symmetric, that is, invariant to relabeling of the options. The online appendix shows that under (5)–(7) the four estimated parameters are point-identified.

Table 2 presents estimates of the Bayesian probit parameters by maximum likelihood, imposing the above restrictions (5)–(7). The estimated utility parameters in table 2 represent the true preferences of a decision maker who maximizes utility under imperfect information. Option *B* has the highest value: $\mu_B = 1.2809$. This means that the Darwinian fitness of male frog *B* is more than one standard deviation above the population mean (better than 90 percent of the population). Option *A* comes in second place, lying 0.1931 standard deviation below the population mean (better than 42 percent of the population). Option *C* is the worst alternative, and its value is 1.0878 standard deviations below the population mean (better than 14 percent of the population). Given this rational benchmark, we may interpret every choice trial in which an inferior option was chosen in the data set as a choice mistake resulting from limited sampling.

The precision parameter $p = 0.0905$ gives a measure of the informational limitation faced by the decision maker. The similarity parameters provide a measure of how much similarity affects ease of comparison for each pair of alternatives. Options *A* and *C* are the most similar, with $\sigma_{AC} = 0.9516$, followed by options *A* and *B*, with $\sigma_{AB} = 0.0484$. The pair *B*, *C* is the least similar with $\sigma_{BC} = 0$. If we interpret the similarity parameter as a measure of "closeness," then *A* and *C* are estimated to be the "closest" pair of options, while *B* and *C* are the "farthest" from each other.

Table 3 shows the choice probabilities generated by the model using the estimated parameter values. The Bayesian probit matches all the qualitative features in the data. In particular, it accommodates the revealed preference relation, the revealed similarity relation, and the two choice frequency reversals that are observed when option *C* is introduced as a choice option and as a phantom option.

B. *Comparison to Random Utility and Rational Inattention*

To compare the performance of my model to the random utility framework, I estimate a five-parameter model of the entire RUM family, using

TABLE 2
POINT ESTIMATES FOR THE BAYESIAN PROBIT MODEL PARAMETERS

	Parameter						
	p	μ_A	μ_B	μ_C	σ_{AB}	σ_{AC}	σ_{BC}
Estimate	.0905	−.1931	1.2809	−1.0878	.0484	.9516	.0000
Standard deviation	.0652	.5239	.6001		.0999		

NOTE.—Point estimates for utility μ_i, similarity σ_{ij}, and precision p in the Bayesian probit model imposing the restrictions in eqq. (5)–(7) and using the full data set of table 1.

TABLE 3
Relative Frequencies of Choice in the Original Data and Estimated Choice Probabilities for the Bayesian Probit, the Random Utility Model, and the Multinomial Logit Model

	Data			BProbit			RUM			Logit		
Choice Problem	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
{*A*, *B*}	.37	.63	. . .	.37	.63	. . .	.48	.52	. . .	.52	.48	. . .
{*A*, *C*}	.84	. . .	.16	.81	. . .	.19	.83	. . .	.17	.77	. . .	.23
{*B*, *C*}	. . .	.69	.31	. . .	.69	.31	. . .	.69	.31	. . .	.75	.25
{*A*, *B*, *C*}	.55	.28	.17	.59	.30	.11	.48	.35	.17	.45	.41	.14
{*A*, *B*, *₵*}	.61	.39	. . .	.63	.37	. . .	.48	.52	. . .	.52	.48	. . .

a characterization of RUM by Block and Marschak (1960) (see the online appendix). The RUM family includes models with independent shocks (such as logit and independent probit) and models in which shocks are allowed to be correlated (such as nested logit and correlated probit). The fit of the RUM specification in table 3 is, by definition, superior to the fit of any of the special cases of the RUM family.

The first three rows of table 3 show that the random utility model accommodates the choice patterns observed in the binary choice data. In particular, the RUM family is able to accommodate the revealed preference and the revealed similarity relations. Unlike the Bayesian probit model, however, the RUM family is unable to accommodate the choice frequency reversals in the data. First, RUM restricts the probability of choosing *A* over *B* to be the same independently of the presence of the phantom option *₵*. Thus, in the RUM model the first row and the last row of table 3 have to be identical. Second, in the RUM model the probability of choosing *A* from {*A*, *B*} cannot increase when the choice option *C* is introduced. 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.

Next, I compare the estimates of my model to estimates of the *rational inattention* model pioneered by Sims (2003). Like the Bayesian probit model, the rational inattention model assumes that the decision maker has a prior over the value of the options, updates her beliefs on the basis of the observation of noisy signals about the values, and chooses the option with the highest expected value given posterior beliefs. The crucial difference from my model is that the decision maker is free to choose an arbitrary distribution of noisy signals, subject to a utility cost of acquiring and processing information. For example, it is feasible for the rationally inattentive decision maker to obtain the Gaussian signals of the Bayesian probit model, but it is also feasible for the decision maker to modify its similarity and precision parameters or to abandon the Gaussian distribution entirely.

TABLE 4
MODEL FIT

Model	k	Log-Like	AIC
BProbit	4	−266.27	540.53
RUM	5	−271.04	552.09
Logit	2	−275.75	555.49
Probit	4	−274.32	556.65

NOTE.—The Bayesian probit provides the best fit according to the AIC. Lower values of AIC are better; k denotes the number of free parameters in each model.

I fit the rational inattention model to the frog mating choice data set under two assumptions: (i) the decision maker's prior is exchangeable, that is, invariant to permutations of the options; and (ii) the utility cost of processing information is proportional to the mutual information measure of Shannon (1948). Under assumptions i and ii, Matejka and McKay (2015) showed that the optimal information acquisition strategy for the rationally inattentive decision maker generates choice probabilities that follow the multinomial logit model.[11]

The last column of table 3 (logit) shows the estimated choice probabilities for the rational inattention model under assumptions i and ii. As in the random utility model, this rational inattention specification fails to account for the two choice frequency reversals observed in the data. In addition, it fails to capture the revealed preference and the revealed similarity relations observed in the binary choice data.

The superiority of the fit of the Bayesian probit model is confirmed by the Akaike information criterion (AIC) shown in table 4. The AIC offers an estimate of the information loss when a given model is used to represent the data (Akaike 1974). The AIC is given by $2k - 2\ln(L)$, 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. Table 4 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 equations (5)–(7) can only improve the fit, which means that the worst possible AIC obtained by a more general version of the Bayesian probit is bounded above by 546.53. Hence, the constraints imposed by equations (5)–(7) do not affect model selection.

[11] Determining the empirical content of the rational inattention model with non-exchangeable priors and other information cost functions is an active area of research. See, e.g., Caplin and Dean (2013, 2015), Matejka and McKay (2015), de Oliveira et al. (2017), Fosgerau, Melo, and Shum (2017), and Hêbert and Woodford (2017) for recent advances in this area.

C. *Out-of-Sample Prediction*

Next, I present five out-of-sample prediction exercises using the published data of Lea and Ryan (2015). In each prediction exercise, I estimate the model excluding the choice trials for a particular choice problem and compare the predicted choice probabilities to the sampled probabilities in that choice problem. I perform this exercise separately for the five choice problems in the experiment.

Table 5 compares the out-of-sample predictions of my model (Bprobit) with the general random utility model (RUM), the multinominal logit model (logit), and the multinomial probit model allowing for correlated shocks (probit). As I explained in Section VII.B, the logit specification is equivalent to the rational inattention model with an exchangeable prior and information costs based on Shannon entropy. The first column of table 5 compares the prediction of each model for choice problem $\{A, B\}$ to the actual data. Each model is estimated excluding all choice trials for choice problem $\{A, B\}$ from the sample. The second column in table 5 performs the same exercise for choice problem $\{A, C\}$, and so on. An asterisk (*) in the entry means that a model does not have enough empirical bite to predict a single choice probability.

The Bayesian probit performs better than the other models in every exercise, except in choice problem $\{A, B\}$. Note that excluding the choice problem $\{A, B\}$ from the sample eliminates any evidence of choice reversals and context effects. In the data, option A is chosen more often than B in every choice problem except $\{A, B\}$. Using the restricted data set, every model including the Bayesian probit estimates that option A has a higher utility than B (see the online appendix) and predicts that A is chosen more often than B in $\{A, B\}$.

TABLE 5
OUT-OF-SAMPLE PREDICTION

	Choice Problem											
	(1)		(2)		(3)		(4)			(5)		
	A	*B*	*A*	*C*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*₵*
Data	.37	.63	.84	.16	.69	.31	.55	.28	.17	.61	.39	. . .
BProbit	.81	.19	.71	.29	.73	.27	.58	.35	.07	.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	. . .

NOTE.—Choice frequencies in the original data vs. choice frequencies predicted by each model in five pseudo out-of-sample exercises. Each exercise excludes all the choice trials for one choice problem from the estimation. An asterisk (*) indicates that the model does not have enough empirical bite to predict a single value. The online appendix contains estimation details.

The Bayesian probit model allows (but does not require) the choice reversal to happen. For the sake of illustrating this point, I provide the local maximum for the likelihood function in the region $\mu_B > \mu_A$. The local maximum is obtained at $p = 0.0985$, $\mu_A = -0.1268$, $\mu_B = 1.1923$, $\mu_C = -1.0654$, $\sigma_{AB} = 0.0579$, $\sigma_{AC} = 0.9421$, and $\sigma_{BC} = 0$ achieving a log likelihood of -188.33. The predicted choice probabilities for $\{A, B\}$ are .38 and .62 for A and B, respectively. While this estimate makes a prediction very close to the actual observed frequencies in $\{A, B\}$, it is not as likely to have generated the restricted data set in which $\{A, B\}$ is excluded (the ratio of likelihoods is $e^{-188.33+187.74} = 0.55$).

VIII. Possible Extensions

I conclude the paper by discussing two possible extensions of the framework.

A. Taste Heterogeneity

In the empirical application in Section VII, I fit a single utility function to the aggregate choices of hundreds of female frogs. Hence, I explained the data as arising from the maximization of a single preference over the options. In my explanation of the data, every female 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 data (as I showed in Sec. II), and it is also not necessary (as table 3 demonstrates).

With human subjects, however, allowing for taste heterogeneity is natural. Analyzing aggregated choices of different individuals may require specifying the model conditional on a vector of individual characteristics, or enriching the model with an additional error term to account for heterogeneous tastes, in addition to perceptual errors. But if multiple choice trials are observed for each individual subject, accounting for heterogeneous tastes can be accomplished simply by fitting the model individually for each subject. This allows different subjects to exhibit different tastes, different levels of risk aversion, and so forth.

Fitting the model to multiple choices by the same subject, however, requires an identifying assumption that signals and posterior beliefs are independent across choice trials. This may be a reasonable assumption in many settings. Take, for example, the experimental work of Soltani et al. (2012) involving lotteries over money shown to subjects on a computer. In these experiments subjects had only a few seconds to contemplate the options before making a choice (4 seconds in choice trials with two options and 8 seconds in choice trials with three options). Each subject participated in multiple sessions and made hundreds of choices. To avoid repeating choice problems, the position of the options on the screen was randomized, and the prizes and probabilities of the options were slightly perturbed in each trial. To fit the model to such individual data, the analyst assumes that, even if the exact same lottery appears on the screen in two different choice trials for the same subject, it is treated by the subject as an independent draw from the prior distribution of utility.

B. Endogenous Precision: Choosing When to Stop Sampling

A second possible extension of this framework is to neuroeconomic applications in which the analyst observes, in addition to choices, the time that it took subjects to reach a decision (see, e.g., Krajbich, Armel, and Rangel 2010). In this case, one may be interested in explaining and predicting the joint distribution of observed choices and time.

An extension of the estimation framework that could be useful in such applications formulates the signals observed by the decision maker as a continuous stochastic process $X : \Omega \times [0, \infty) \to \mathbb{R}^n$, where Ω is the underlying probability space. The signal process starts from $X(0) = 0$ almost surely and follows a Brownian motion with drift given by

$$dX(t) = \mu\, dt + \Lambda\, dW(t), \tag{8}$$

where the constant drift vector $\mu = (\mu_1, \dots, \mu_n)$ represents the true utility of the options, Λ is a constant $n \times n$ matrix with full rank, and $W(t) = (W_1(t), \dots, W_n(t))$ is a Wiener process.[12] A decision maker who observes this continuous process up to time T has the same information as a decision maker who observes a single realization of the signals in the static formulation with precision $p = T$.[13]

Suppose that the time spent contemplating the options before reaching a decision is chosen endogenously by the decision maker, who solves

[12] Drift-diffusion models like (8) are used in neuroscience to represent the noisy process by which the human brain perceives the value of choice options over time. These models typically focus on binary choice and assume uncorrelated signals (Ratcliff 1978; Busemeyer and Townsend 1993; Woodford 2014). Fehr and Rangel (2011) provide an overview of this literature.

[13] The continuous learning process formulation provides a justification for the assumption that the decision maker knows the value of the correlation parameters σ_{ij}. Since the continuous random process accumulates quadratic variation, the decision maker can perfectly estimate the parameters σ_{ij} after an arbitrarily small time interval (Shreve 2004). Hence, the value of the similarity parameters σ is instantly learned, while the value of the utility parameters μ is learned gradually over time.

an optimal stopping problem with costly information acquisition. The framework in this paper corresponds to a special family of contemplation cost functions, where cost is constant and equal to zero up to a fixed level of precision $t = p$, jumping to plus infinity afterward. An alternative specification with the same number of free parameters may impose a linear fixed cost $c > 0$ per unit of time, obtaining the distribution of stopping times endogenously. For an example of extending the framework in this direction, the paper by Fudenberg, Strack, and Strzalecki (2018) studies the optimal stopping problem in the special case of binary choice and independent signal distributions.

Appendix

Proofs

Proof of Lemma 1

In the binary choice problem $\{i, j\}$ with no phantom options the decision maker observes two signals (X_i, X_j). The vector of posterior mean beliefs (m_i, m_j) for options i, j has a joint Gaussian distribution. The difference $m_j - m_i$ is therefore Gaussian. The mean and variance of $m_j - m_i$ are given, respectively, by

$$\left(\frac{t(\mu_j - \mu_i)}{t + 1 - \sigma_{ij}}, \frac{2t(1 - \sigma_{ij})}{(t + 1 - \sigma_{ij})^2}\right),$$

and therefore, we have

$$\begin{aligned} P_{\mu,\sigma,p}(i, \{i, j\}) &= \mathbb{P}\{m_j - m_i < 0\} \\ &= \mathbb{P}\left\{\frac{m_j - m_i - \mathbb{E}[m_j - m_i]}{\sqrt{\mathrm{Var}[m_j - m_i]}} < -\frac{\mathbb{E}[m_j - m_i]}{\sqrt{\mathrm{Var}[m_j - m_i]}}\right\} \\ &= \Phi\left(\frac{\sqrt{t}(\mu_i - \mu_j)}{\sqrt{2(1 - \sigma_{ij})}}\right) \\ &= \mathbb{P}\{X_j - X_i < 0\} \\ &= \ddot{P}_{\mu,\sigma,p}(i, \{i, j\}), \end{aligned}$$

which proves part i.

To prove part ii, suppose $\sigma_{ij} = c_1$ is constant for all i, j. Fix a choice problem (N, M) and enumerate its options $N \cup M = \{1, \ldots, n\}$. Posterior beliefs are

$$(m_1, \ldots, m_n) = \mathbb{E}[\mu | X] = \left[I + p^{-1}\Sigma_n\right]^{-1}(X_1, \ldots, X_n).$$

By assumption the diagonal entries in Σ_n are all equal to one, while the off-diagonal elements are all equal to c_1. This implies that the matrix $[I + p^{-1}\Sigma_n]^{-1}$ is symmetric with all diagonal entries equal to a constant c_2 and all off-diagonal entries equal to a constant c_3. Thus $m_i = c_2 X_i + c_3 \Sigma_{k \neq i} X_k$ for all i, and for any possible signal realization we have $m_i > m_j$ if and only if $X_i > X_j$.

To prove part iii, fix again the set of options in a choice problem $N \cup M = \{1, \dots, n\}$ and let $m(p) = (m_1(p), \dots, m_n(p))$ be the corresponding vector of posterior mean beliefs, making explicit the dependence on precision p. Using the matrix identity $(I + A)^{-1} = I - (I + A)^{-1}A$, we obtain

$$\sqrt{p}\,\mu - \mathbb{E}[\sqrt{p}\,m(p)] = \frac{1}{\sqrt{p}}[I + p^{-1}\Sigma]^{-1}\Sigma\mu.$$

Taking the limit we obtain $\lim_{p\to\infty}|\mathbb{E}[\sqrt{p}\,m(p)] - \sqrt{p}\,\mu| = 0$. Moreover,

$$\lim_{p\to\infty}\mathrm{Var}[\sqrt{p}\,m(p)] = \lim_{p\to\infty}[I + p^{-1}\Sigma]^{-1}\Sigma[I + p^{-1}\Sigma]^{-1} = \Sigma.$$

Hence, the mean and covariance matrix of the Gaussian vector $\sqrt{p}\,m(p)$ remain arbitrarily close to the mean and covariance matrix of $\sqrt{p}\,X(p)$, respectively, for p sufficiently large. Since for all $p > 0$ we have

$$\begin{aligned} P_{\mu,\sigma,p}(i, \{1, \dots, n\}) &= \mathbb{P}\left(\cap_{j\neq i}\left\{ m_i(p) > m_j(p)\right\}\right) \\ &= \mathbb{P}\left(\cap_{j\neq i}\left\{ \sqrt{p}\,m_i(p) > \sqrt{p}\,m_j(p)\right\}\right), \\ \ddot{P}_{\mu,\sigma,p}(i, \{1, \dots, n\}) &= \mathbb{P}\left(\cap_{j\neq i}\left\{ X_i(p) > X_j(p)\right\}\right) \\ &= \mathbb{P}\left(\cap_{j\neq i}\left\{ \sqrt{p}\,X_i(p) > \sqrt{p}\,X_j(p)\right\}\right), \end{aligned}$$

part iii follows by continuity. QED

Proof of Proposition 1

Suppose μ_i are independently and identically distributed $\mathcal{N}(m, s^2)$ according to the decision maker's prior. Fix a menu of alternatives labeled $\{1, \dots, n\}$. Applying Bayes's rule (see, e.g., DeGroot 1970), the posterior is Gaussian with mean vector

$$(m, \dots, m) + s^{-1}[s^{-1}I + p^{-1}\Sigma]^{-1}[X - (m, \dots, m)].$$

Choice probabilities remain the same if we subtract the constant m from each coordinate and rescale the vector by $\sqrt{s}$, obtaining

$$s^{-1}[s^{-1}I + p^{-1}\Sigma]^{-1}\sqrt{s}[X - (m, \dots, m)].$$

Conditional on μ (from the point of view of the analyst), this vector is equal to

$$[ss^{-1}I + sp^{-1}\Sigma]^{-1}\{\sqrt{s}[\mu - (m, \dots, m)] + \sqrt{s}\,\varepsilon\},$$

which is the posterior mean vector when the prior is $\mathcal{N}(0, 1)$, utility is $\tilde{\mu} = \sqrt{s}[\mu - (m, \dots, m)]$, the correlation matrix is $\tilde{\Sigma} = \Sigma$, and precision is $\tilde{p} = p/s$. QED

Proof of Proposition 2

The binary choice formula (4) implies that $P_{\mu,\sigma,p}(i, \{i, j\}) > .5$ if and only if $\Phi^{-1}(P_{\mu,\sigma,p}(i, \{i, j\})) > \Phi^{-1}(.5) = 0$ if and only if $\mu_i - \mu_j > 0$, proving part i.

Suppose that the pair $\{i, j\}$ is revealed more similar than $\{k, \ell\}$. Then $\{i, j\}$ is easier to compare than $\{k, \ell\}$, and relabeling the options in each pair if necessary,

$k \succsim i \succ j \succsim \ell$. By part i above, we have $\mu_k \geq \mu_i > \mu_j \geq \mu_\ell$, which implies $\mu_k - \mu_\ell \geq \mu_i - \mu_j > 0$. Since $\{i, j\}$ is easier to compare than $\{k, \ell\}$, the binary choice formula (4) implies

$$\frac{\mu_i - \mu_j}{\sqrt{1 - \sigma_{ij}}} > \frac{\mu_k - \mu_\ell}{\sqrt{1 - \sigma_{kl}}} \geq \frac{\mu_i - \mu_j}{\sqrt{1 - \sigma_{kl}}},$$

which in turn implies $\sigma_{ij} > \sigma_{k\ell}$, proving part ii. QED

Proof of Proposition 3

Lemma 1, part i, shows that the restriction of $P_{\mu,\sigma,p}$ to binary choices is observationally equivalent to a multinomial probit, which satisfies MT (see Halff 1976).

In every choice problem that appears in the statement of MT-P, the union of the set of choice options and the set of phantom options is equal to $M \cup \{i, j, k\}$. This implies that the set of signals is the same and posterior beliefs have the same Gaussian distribution. Binary choices are again observationally equivalent to a multinomial probit, and MT-P then follows from MT.

It remains to show that BM-P holds. Block and Marschak (1960) showed that the stochastic choice function of any random utility model over a finite collection of choice objects $N' = \{1, \ldots, n\}$ satisfies a set of inequalities,

$$(\forall\ i \in N \subseteq N') \sum_{M : N \subseteq M \subseteq N'} (-1)^{|M \setminus N|} P(i, M) \geq 0. \tag{BM}$$

The statement of BM-P involves only choice problems of the form $(M, N' \setminus M)$. Across all such choice problems, the set of signals observed by the decision maker is equal to $(X_i)_{i \in N'}$. This implies that posterior beliefs have the same distribution across all such choice problems. Hence, across these choice problems, the Bayesian probit is equivalent to a random utility model, in which posterior beliefs play the role of random utilities. Then BM-P follows from BM above. QED

Proof of Proposition 4

LEMMA 2. The limit of $\ddot{P}_{\mu,\sigma,p}(1, \{1, 2, 3\})$ when $p \to 0$ is equal to

$$\frac{1}{4} + \frac{1}{2\pi} \arctan\left(\frac{1 + \sigma_{23} - \sigma_{12} - \sigma_{13}}{\sqrt{4(1 - \sigma_{12})(1 - \sigma_{13}) - (1 + \sigma_{23} - \sigma_{12} - \sigma_{13})^2}} \right).$$

Proof. Let $X = (X_1, X_2, X_3)'$ be the vector of Gaussian signals with mean μ and covariance matrix $p^{-1}\Sigma$ and let L be the 2×3 matrix given by

$$L = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \tag{A1}$$

so that

$$\ddot{P}_{\mu,\sigma,p}(1, \{1, 2, 3\}) = \mathbb{P}\{LX \leq 0\} = \mathbb{P}\{\sqrt{p} L\varepsilon \leq -\sqrt{p} L\mu\}.$$

Let Λ be the unique 2×2 Cholesky decomposition matrix such that $\Lambda\Lambda' = L\Sigma L'$. If $Z = (Z_1, Z_2)$ is a standard Gaussian vector, then ΛZ and $\sqrt{p}L\varepsilon$ have the same Gaussian distribution. This implies

$$\lim_{p\to 0} \ddot{P}_{\mu,\sigma,p}(1, \{1,2,3\})$$

$$= \mathbb{P}\left\{ Z_1 \leq 0 \text{ and } Z_2 \leq -Z_1 \frac{1 + \sigma_{23} - \sigma_{12} - \sigma_{13}}{\sqrt{4(1-\sigma_{12})(1-\sigma_{13}) - (1 + \sigma_{23} - \sigma_{12} - \sigma_{13})^2}} \right\},$$

which is the probability that (Z_1, Z_2) lies in a cone in $\mathbb{R}^2$. The result then follows from the circular symmetry of the standard Gaussian distribution. QED

Since by assumption $\mu_1 = \mu_2 = \mu_3$, we have

$$\begin{aligned}\ddot{P}_{\mu,\sigma,p}(i, \{1,2,3\}) &= \mathbb{P}\{X_i(p) \geq X_j(p) \text{ for all } j\} \\ &= \mathbb{P}\{\sqrt{p}[X_i(p) - \mu_i] \geq \sqrt{p}[X_j(p) - \mu_j] \text{ for all } j\}.\end{aligned}$$

For every $p > 0$, the vector $\sqrt{p}[X(p) - \mu]$ is jointly Gaussian with mean vector zero and covariance matrix Σ. Hence, for all $p, p' > 0$,

$$\ddot{P}_{\mu,\sigma,p}(i, \{1,2,3\}) = \ddot{P}_{\mu,\sigma,p'}(i, \{1,2,3\}) = \ddot{P}_{\mu,\sigma,0}(i, \{1,2,3\}),$$

where

$$\ddot{P}_{\mu,\sigma,0}(i, \{1,2,3\}) = \lim\nolimits_{p\to 0}\ddot{P}_{\mu,\sigma,p}(i, \{1,2,3\}).$$

By the assumption $\sigma_{23} > \sigma_{13}$ and lemma 2, we obtain, for all $p > 0$,

$$\begin{aligned}&\ddot{P}_{\mu,\sigma,p}(1, \{1,2,3\}) - \ddot{P}_{\mu,\sigma,p}(2, \{1,2,3\}) \\ &= \ddot{P}_{\mu,\sigma,0}(1, \{1,2,3\}) - \ddot{P}_{\mu,\sigma,0}(2, \{1,2,3\}) > 0.\end{aligned}$$

Finally, lemma 1, part iii, implies $P_{\mu,\sigma,p}(1, \{1,2,3\}) > P_{\mu,\sigma,p}(2, \{1,2,3\})$ for all p sufficiently large. QED

Proof of Proposition 5

LEMMA 3. The limit of $P_{\mu,\sigma,p}(1, \{1,2,3\})$ as $p \to 0$ is equal to

$$\frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{(1+\sigma_{12})(1+\sigma_{13}) - \sigma_{23}(1 + \sigma_{12} + \sigma_{13} + \sigma_{23})}{\sqrt{(3 + 2\sigma_{12} + 2\sigma_{13} + 2\sigma_{23})(1 + 2\sigma_{12}\sigma_{13}\sigma_{23} - \sigma_{12}^2 - \sigma_{13}^2 - \sigma_{23}^2)}}\right).$$

Proof. Let $X = (X_1, X_2, X_3)'$ be the vector of Gaussian signals with mean μ and covariance matrix $p^{-1}\Sigma$. Let $m(p) = [I + p^{-1}\Sigma]^{-1}X$ be the vector of posterior beliefs for a given $p > 0$ and let L be the 2×3 matrix given by equation (A1) so that $P_{\mu,\sigma,p}(1, \{1,2,3\}) = \mathbb{P}\{Lm(p) \leq 0\}$. Let $Z = (Z_1, Z_2)$ be a standard Gaussian vector, and let $\Lambda(p)$ be the unique 2×2 Cholesky decomposition matrix such that

$$\Lambda(p)\Lambda(p)' = L[I + p^{-1}\Sigma]^{-1}p^{-1}\Sigma[I + p^{-1}\Sigma]^{-1}L'.$$

Then $Lm(p)$ and $L[I + p^{-1}\Sigma]^{-1}\mu + \Lambda(p)Z$ have the same Gaussian distribution for every $p > 0$. This implies that the limit of $P_{\mu,\sigma,p}(1, \{1, 2, 3\})$ when $p \to 0$ is equal to

$$\mathbb{P}\left\{Z_1 \leq 0 \text{ and } Z_2 \leq -Z_1 \frac{(1+\sigma_{12})(1+\sigma_{13}) - \sigma_{23}(1+\sigma_{12}+\sigma_{13}+\sigma_{23})}{\sqrt{(3+2\sigma_{12}+2\sigma_{13}+2\sigma_{23})(1+2\sigma_{12}\sigma_{13}\sigma_{23}-\sigma_{12}^2-\sigma_{13}^2-\sigma_{23}^2)}}\right\},$$

which is the probability that (Z_1, Z_2) lies in a cone in $\mathbb{R}^2$. The result then follows from the circular symmetry of the standard Gaussian distribution. QED

Suppose $\sigma_{12} > -1/2$ and $\sigma_{23} > \sigma_{13}$. Lemma 3 above provides the probability that option 1 is chosen from {1, 2, 3} in the limit as $p \to 0$. The choice probability for option 2 is obtained by relabeling the options in lemma 3. Since the arctan function is strictly increasing, the probability of option 1 is strictly smaller than the probability of option 2 in the limit as $p \to 0$ if and only if

$$(\sigma_{23} - \sigma_{13})(2 + 2\sigma_{12} + \sigma_{13} + \sigma_{23}) > 0.$$

The first term in parentheses must be strictly positive, since $\sigma_{23} > \sigma_{13}$. The second term in parentheses is also strictly positive; it follows from $(1 + 2\sigma_{12}\sigma_{13}\sigma_{23} - \sigma_{12}^2 - \sigma_{13}^2 - \sigma_{23}^2) = \det\Sigma > 0$ and $\sigma_{12} > -1/2$. Therefore, we have

$$\lim_{p\to 0}\frac{P_{\mu,\sigma,p}(1, \{1, 2, 3\})}{P_{\mu,\sigma,p}(2, \{1, 2, 3\})} < 1.$$

On the other hand, lemma 1, part i, implies that the ratio $P_{\mu,\sigma,p}(1, \{1, 2\})/P_{\mu,\sigma,p}(2, \{1, 2\})$ goes to one as p shrinks to zero. Thus, for all sufficiently small values of $p > 0$, introducing option 3 hurts option 1 proportionally more than option 2. QED

Proof of Proposition 6

The determinant of the covariance matrix of signals corresponding to the menu of options {1, 2, 3} is strictly positive if and only if $1 + 2\sigma_{12}\sigma_{13}\sigma_{23} - \sigma_{12}^2 - \sigma_{13}^2 - \sigma_{23}^2 > 0$. Thus, for any fixed values of σ_{12} and σ_{13}, the value of σ_{23} is bounded above:

$$\sigma_{23} < \sigma_{12}\sigma_{13} + \sqrt{(1 - \sigma_{12}^2)(1 - \sigma_{13}^2)}.$$

For any fixed σ_{12}, option 3 will satisfy $\sigma_{23} > \sigma_{12}$ and $\sigma_{23} > \sigma_{13}$ if and only if the upper bound for σ_{23} is strictly larger than σ_{12} and σ_{13}. Hence, for any fixed value of σ_{12}, the correlation σ_{13} must be in the following range:

$$\sigma_{13} \in \begin{cases} \left(-1, \sqrt{1/2 + \sigma_{12}/2}\right) & \text{if } \sigma_{12} \leq 0 \\ \left(2\sigma_{12}^2 - 1, \sqrt{1/2 + \sigma_{12}/2}\right) & \text{if } \sigma_{12} > 0. \end{cases}$$

It can be shown that, for every fixed pair $(\sigma_{12}, \sigma_{13})$ with $-1 < \sigma_{12} < 1$ and σ_{13} in the range above, in the limit as σ_{23} increases toward its upper bound,

$$\frac{(1+\sigma_{12})(1+\sigma_{13})-\sigma_{23}(1+\sigma_{12}+\sigma_{13}+\sigma_{23})}{\sqrt{(3+2\sigma_{12}+2\sigma_{13}+2\sigma_{23})(1+2\sigma_{12}\sigma_{13}\sigma_{23}-\sigma_{12}^2-\sigma_{13}^2-\sigma_{23}^2)}}\to-\infty.$$

This holds because the numerator goes to a strictly negative value, while the strictly positive denominator goes to zero. The result follows from lemma 3. QED

Proof of Proposition 7

Option 2 is chosen from {1, 2, 3} when both coordinates of $(m_1(p)-m_2(p), m_3(p)-m_2(p))$ are negative. The mean of the first coordinate is given by

$$\begin{aligned}\mathbb{E}[m_1(p)-m_2(p)] = {} & p\{\mu_3(1+p+\sigma_{12})(\sigma_{23}-\sigma_{13})\\ & +\mu_2[-(1+p)(1+p+\sigma_{12})+\sigma_{23}\sigma_{13}+\sigma_{13}^2]\\ & +\mu_1[1+\sigma_{12}+p(2+p+\sigma_{12})-\sigma_{23}(\sigma_{23}+\sigma_{13})]\}/\\ & \left\{(1+p)[1+p(2+p)-\sigma_{12}^2-\sigma_{23}^2-\sigma_{13}^2]+2\sigma_{12}\sigma_{23}\sigma_{13}\right\},\end{aligned}$$

and the mean of the second coordinate is given by

$$\begin{aligned}\mathbb{E}[m_3(p)-m_2(p)] = {} & p\{\mu_1(1+p+\sigma_{23})(\sigma_{12}-\sigma_{13})\\ & +\mu_2[-(1+p)(1+p+\sigma_{23})+\sigma_{12}\sigma_{13}+\sigma_{13}^2]\\ & +\mu_3[1+\sigma_{23}+p(2+p+\sigma_{23})-\sigma_{12}(\sigma_{12}+\sigma_{13})]\}/\\ & \left\{(1+p)[1+p(2+p)-\sigma_{12}^2-\sigma_{13}^2-\sigma_{23}^2]+2\sigma_{12}\sigma_{23}\sigma_{13}\right\}.\end{aligned}$$

The denominator is equal in both expressions and can be written as

$$p[p(3+p)+3-\sigma_{12}^2-\sigma_{13}^2-\sigma_{23}^2]+[1+2\sigma_{12}\sigma_{23}\sigma_{13}-\sigma_{12}^2-\sigma_{13}^2-\sigma_{23}^2],$$

which is strictly positive for all $p>0$, since $\sigma_{ij}^2<1$ and $\det\Sigma>0$.

In both numerators above, the expression multiplying the coefficient μ_3 is strictly positive. In the first case, $1+p+\sigma_{12}>0$ is always strictly positive and $\sigma_{23}>\sigma_{13}$ by assumption. In the second case, the expression multiplying μ_3 can be written as

$$[1-\sigma_{12}^2]+[p(2+\sigma_{23}+p)]+[\sigma_{23}-\sigma_{12}\sigma_{13}],$$

where each expression in brackets is clearly positive. Therefore, for any fixed $p>0$, the expectations $\mathbb{E}[m_1(t)-m_2(t)]$ and $\mathbb{E}[m_3(t)-m_2(t)]$ can be made arbitrarily negative by taking $\mu_3<0$ and sufficiently large in absolute value.

Since by assumption $\mu_1=\mu_2$, both $P_{\mu,\sigma,p}(1,\{1,2,3\})$ and $P_{\mu,\sigma,p}(2,\{1,2,3\})$ converge to 1/2 when p goes to infinity. The covariance matrix

$$\mathrm{Var}[m(p)]=[I+p^{-1}\Sigma]^{-1}p^{-1}\Sigma[I+p^{-1}\Sigma]^{-1}$$

does not depend on μ. Hence, increasing the absolute value of the negative parameter μ_3 does not change $\mathrm{Var}[m(p)]$ for any p, while decreasing $\mathbb{E}[m_1(p)-m_2(p)]$ and $\mathbb{E}[m_3(p)-m_2(p)]$ for every $p>0$, and therefore increases $P_{\mu,\sigma,p}(2,\{1,2,3\})$ for every $p>0$. Moreover, for any given fixed $p>0$, $P_{\mu,\sigma,p}(2,\{1,2,3\})$ can be made arbitrarily close to 1 by taking μ_3 sufficiently negative. This guarantees that we can have a violation of monotonicity for any arbitrarily low level of precision $p>0$. Moreover, since $\mathbb{E}[m_1(p) - m_2(p)]$ converges to 0 from below as p goes to infinity, $P_{\mu,\sigma,p}(2, \{1, 2, 3\})$ will converge to 1/2 from above, while $P_{\mu,\sigma,p}(1, \{1, 2, 3\})$ will converge to 1/2 from below. QED

Proof of Proposition 8

As in the proof of proposition 5, I use lemma 3 to show that $P_{\mu,\sigma,0}(2, \{1, 2, 3\}) > P_{\mu,\sigma,0}(1, \{1, 2, 3\})$ if and only if $(\sigma_{23} - \sigma_{13})(2 + 2\sigma_{12} + \sigma_{13} + \sigma_{23}) > 0$. Likewise, $P_{\mu,\sigma,0}(2, \{1, 2, 3\}) > P_{\mu,\sigma,0}(3, \{1, 2, 3\})$ if and only if $(\sigma_{12} - \sigma_{13})(2 + 2\sigma_{23} + \sigma_{12} + \sigma_{13}) > 0$. Since by assumption $\sigma_{12} = \sigma_{23} > \sigma_{13}$, we have both $\sigma_{23} - \sigma_{13} > 0$ and $\sigma_{12} - \sigma_{13} > 0$. The assumption $\sigma_{12} = \sigma_{23} > \sigma_{13}$ and the strictly positive determinant of Σ together imply $(2 + 2\sigma_{12} + \sigma_{13} + \sigma_{23}) > 0$ and $(2 + 2\sigma_{23} + \sigma_{12} + \sigma_{13}) > 0$. By continuity, we have $P_{\mu,\sigma,p}(1, \{1, 2, 3\}) < P_{\mu,\sigma,p}(2, \{1, 2, 3\})$ and $P_{\mu,\sigma,p}(3, \{1, 2, 3\}) < P_{\mu,\sigma,p}(2, \{1, 2, 3\})$ for all $p > 0$ sufficiently small. QED

References

Akaike, Hirotugu. 1974. "A New Look at the Statistical Model Identification." *IEEE Trans. Automatic Control* 19 (6): 716–23.

Bateson, Melissa, Susan D. Healy, and T. Andrew Hurly. 2002. "Irrational Choices in Hummingbird Foraging Behaviour." *Animal Behaviour* 63 (3): 587–96.

Block, H. D., and Jacob Marschak. 1960. "Random Orderings and Stochastic Theories of Responses." In *Contributions to Probability and Statistics*, edited by I. Olkin et al., 97–132. Stanford, CA: Stanford Univ. Press.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2013. "Salience and Consumer Choice." *J.P.E.* 121 (5): 803–43.

Busemeyer, Jerome R., and James T. Townsend. 1993. "Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment." *Psychological Rev.* 100 (3): 432–59.

Caplin, Andrew, and Mark Dean. 2013. "Behavioral Implications of Rational Inattention with Shannon Entropy." Technical report, NBER, Cambridge, MA.

———. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *A.E.R.* 105 (7): 2183–2203.

Caplin, Andrew, Mark Dean, and Daniel Martin. 2011. "Search and Satisficing." *A.E.R.* 101 (7): 2899–2922.

de Clippel, Geoffroy, and Kfir Eliaz. 2012. "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects." *Theoretical Econ.* 7 (1): 125–62.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: Wiley-Interscience.

de Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. "Rationally Inattentive Preferences and Hidden Information Costs." *Theoretical Econ.* 12 (2): 621–54.

Domencich, Thomas A., and Daniel McFadden. 1975. *Urban Travel Demand: A Behavioral Analysis*. Contributions to Economic Analysis, vol. 93. Amsterdam: North-Holland.

Fehr, Ernst, and Antonio Rangel. 2011. "Neuroeconomic Foundations of Economic Choice—Recent Advances." *J. Econ. Perspectives* 25 (4): 3–30.

Fosgerau, Mogens, Emerson Melo, and Matthew Shum. 2017. "Discrete Choice and Rational Inattention: A General Equivalence Result." Working paper, Soc. Sci. Res. Network. https://ssrn.com/abstract=2889048.

Frederick, Shane, Leonard Lee, and Ernest Baskin. 2014. "The Limits of Attraction." *J. Marketing Res.* 51 (4): 487–507.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015. "Stochastic Choice and Revealed Perturbed Utility." *Econometrica* 83 (November): 2371–2409.

Fudenberg, Drew, Philipp Strack, and Tomasz Strzalecki. 2018. "Speed, Accuracy, and the Optimal Timing of Choices." *A.E.R.* 108 (12): 3651–84.

Gul, Faruk, Paulo Natenzon, and Wolfgang Pesendorfer. 2014. "Random Choice as Behavioral Optimization." *Econometrica* 82 (September): 1873–1912.

Halff, Henry M. 1976. "Choice Theories for Differentially Comparable Alternatives." *J. Math. Psychology* 14 (3): 244–46.

Harbaugh, Rick, John Maxwell, and Kelly Shue. 2016. "Consistent Good News and Inconsistent Bad News." Working paper. https://kelley.iu.edu/riharbau/.

Hausman, Jerry A., and David A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica* 46 (2): 403–26.

Hêbert, Benjamin, and Michael Woodford. 2017. "Rational Inattention with Sequential Information Sampling." Working Paper no. 23787, NBER, Cambridge, MA.

Huber, Joel, John W. Payne, and Christopher P. Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *J. Consumer Res.* 9 (1): 90–98.

———. 2014. "Let's Be Honest about the Attraction Effect." *J. Marketing Res.* 51 (4): 520–25.

Kamenica, Emir. 2008. "Contextual Inference in Markets: On the Informational Content of Product Lines." *A.E.R.* 98 (5): 2127–49.

Krajbich, Ian, Carrie Armel, and Antonio Rangel. 2010. "Visual Fixations and the Computation and Comparison of Value in Simple Choice." *Nature Neuroscience* 13 (10): 1292–98.

Latty, Tanya, and Madeleine Beekman. 2011. "Irrational Decision-Making in an Amoeboid Organism: Transitivity and Context-Dependent Preferences." *Proc. Royal Soc. London, B: Biological Sci.* 278 (1703): 307–12.

Lea, Amanda M., and Michael J. Ryan. 2015. "Irrationality in Mate Choice Revealed by Túngara Frogs." *Science* 349 (6251): 964–66.

Lu, Jay. 2016. "Random Choice and Private Information." *Econometrica* 84 (November): 1983–2027.

Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica* 82 (May): 1153–76.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay. 2012. "Revealed Attention." *A.E.R.* 102 (5): 2183–2205.

Matejka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *A.E.R.* 105 (1): 272–98.

Ok, E. A., P. Ortoleva, and G. Riella. 2015. "Revealed (P)reference Theory." *A.E.R.* 105 (January): 299–321.

Parrish, Audrey E., Theodore A. Evans, and Michael J. Beran. 2015. "Rhesus Macaques (*Macaca mulatta*) Exhibit the Decoy Effect in a Perceptual Discrimination Task." *Attention, Perception, and Psychophysics* 77 (5): 1715–25.

Ratcliff, Roger. 1978. "A Theory of Memory Retrieval." *Psychological Rev.* 85 (2): 59–108.

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

Shafir, Sharoni, Tom A. Waite, and Brian H. Smith. 2002. "Context-Dependent Violations of Rational Choice in Honeybees (*Apis mellifera*) and Gray Jays (*Perisoreus canadensis*)." *Behavioral Ecology and Sociobiology* 51 (2): 180–87.

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Tech. J.* 27 (July): 379–423.
Shreve, Steven E. 2004. *Stochastic Calculus for Finance*. Vol. 2: *Continuous-Time Models*. New York: Springer.
Simonson, Itamar. 1989. "Choice Based on Reasons: The Case of Attraction and Compromise Effects." *J. Consumer Res.* 16 (2): 158–74.
———. 2014. "Vices and Virtues of Misguided Replications: The Case of Asymmetric Dominance." *J. Marketing Res.* 51 (4): 514–19.
Sims, Christopher A. 2003. "Implications of Rational Inattention." *J. Monetary Econ.* 50 (3): 665–90.
Soltani, Alireza, Benedetto De Martino, and Colin Camerer. 2012. "A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence." *PLoS Computational Biology* 8 (7): e1002607.
Thurstone, L. L. 1927a. "A Law of Comparative Judgment." *Psychological Rev.* 34 (4): 273.
———. 1927b. "Psychophysical Analysis." *American J. Psychology* 38 (3): 368–89.
Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge: Cambridge Univ. Press.
Tversky, Amos. 1972. "Elimination by Aspects: A Theory of Choice." *Psychological Rev.* 79 (4): 281–99.
Tversky, Amos, and J. Edward Russo. 1969. "Substitutability and Similarity in Binary Choices." *J. Math. Psychology* 6 (1): 1–12.
Tversky, Amos, and Itamar Simonson. 1993. "Context-Dependent Preferences." *Management Sci.* 39 (10): 1179–89.
Weibull, Jörgen, Lars-Göran Mattsson, and Mark Voorneveld. 2007. "Better May Be Worse: Some Monotonicity Results and Paradoxes in Discrete Choice under Uncertainty." *Theory and Decision* 63:121–51.
Woodford, Michael. 2014. "Stochastic Choice: An Optimizing Neuroeconomic Model." *A.E.R.* 104 (5): 495–500.
Yang, Sybil, and Michael Lynn. 2014. "More Evidence Challenging the Robustness and Usefulness of the Attraction Effect." *J. Marketing Res.* 51 (4): 508–13.