Submitted to *Management Science*
manuscript

Bayesian Contextual Choices under Imperfect Perception of Attributes

Junnan He
Department of Economics, Sciences Po, 28 Rue des Saints-Pères, 75007 Paris. junnan.he@sciencespo.fr

Classical theories show that choices can be represented by a stable utility function when they satisfy consistency axioms such as transitivity and the independence of irrelevant alternatives. Empirical choice data, however, display several contextual choice effects that violate these axioms. We study a choice model through a fixed underlying utility function and explain contextual choices with a novel informational friction: the agent's perception of the options is affected by attribute-specific noise. Under this friction, the agent learns useful information when she sees more options. Therefore, the agent chooses contextually, exhibiting intransitivity, joint-separate evaluation reversal, the compromise effect, the phantom decoy effect, the attraction effect, and the similarity effect. Because the noise is attribute-specific and common across alternatives, the classical axioms hold when the alternatives dominate one another in attributes.

Key words: compromise effect, context effect, imperfect perception, intransitive choices, joint-separate evaluation reversal, phantom decoy effect, stable preferences.

History: First version 2016, current version 2021

1. Introduction

Modeling consumer choices as the maximization of a stable utility function has practical advantages in the economic analysis of policy interventions and consumer welfare. Classical theories show that when choices satisfy certain consistency axioms, e.g., transitivity and independence of contexts, they can be represented by a stable utility function. However, empirical research has long found that choices often violate these axioms in certain choice problems.[1] In this paper, we provide a natural setting that uses a stable utility function to accommodate several well-documented contextual choice effects. Through our model, a stable and well-defined utility function can still be recovered from such contextual choice data.

By contextual effects, we mean the following type of observation. For two objects $\mathbf{x}$ and $\mathbf{y}$, their observed choice probabilities or reported evaluations differ across decision problems in a way that implies their relative

[1] For example, intransitivity was spotted as early as Tversky (1969), and some recent evidence is surveyed in Rieskamp et al. (2006). Empirical studies on other aspects of contextual dependence include Simonson (1989), Pratkanis and Farquhar (1992), Huber et al. (1982), and Hsee (1996).

value depends on other competing alternatives. Well-known examples include intransitivity, joint-separate evaluation reversals (henceforth j-s reversals), and various violations of the independence of irrelevant alternatives (henceforth IIA). The compromise effect in Simonson (1989) is an example of an IIA violation. In the experiments, the participants face two choice problems. One has only two options $\mathbf{x}$ and $\mathbf{y}$, and the other includes a third option $\mathbf{z}$. The attributes of the three are such that $\mathbf{x}$ is the "middle option" between $\mathbf{y}$ and $\mathbf{z}$. Empirically, including option $\mathbf{z}$ can reverse the relative choice frequency between $\mathbf{x}$ and $\mathbf{y}$, even though $\mathbf{z}$ itself is rarely chosen.

To systematically model these observations, we assume a novel informational friction under which a decision maker maximizes a *stable* preference. We show how such an informational friction induces the aforementioned contextual effects for correctly designed choice problems. We unify three types of IIA violations — the compromise effect, the phantom decoy effect, and the attraction effect — by defining a comparative static called the *decoy choice pattern* and show in Theorem 1 that this pattern is predicted by our model. Compared with other models with high explanatory power, ours is parsimonious in the sense that several contextual effects can be explained within one simple parametric setting. In addition to explaining contextual effects, our model also possesses desirable regularities. For instance, it predicts that the classical consistency axioms hold for the class of choice problems with a dominating alternative in attributes (Theorem 2).

Our novel friction assumes that attribute perception is noisy. An object $\mathbf{x}$ has a vector of unobserved hedonic (or sensory) attributes $\mathbf{x}^*$ that enters the agent's utility function. The agent can only observe its noisy signal $X|\mathbf{x}^*$. Such kinds of noisy perception are ubiquitous among ordinary choice problems, including problems in which the objects' characteristics are measured and displayed in scientific units.[2] One way to conceptualize this is to think that $\mathbf{x}^*$ and X are both defined in the space of hedonic attributes — the domain of the utility function. Being in the same domain, they are directly comparable. The label information, on the other hand, lives in a different space (of verbal or numeric descriptions). The unobserved attribute $\mathbf{x}^*$ is first mapped to the space of label information through relevant measurements. The label information is then mapped back to the hedonic domain through interpretation. The output of the interpretation mapping is the signal X.

$$\text{unobserved attributes } \mathbf{x}^* \xrightarrow{\text{relevant measurements}} \text{label information} \xrightarrow{\text{interpretations}} \text{observed signal } \mathbf{X}.$$

As argued in Ariely et al. (2003) and Kamenica (2008), scientific measurements and numerical information are hard to interpret. Such information is usually not natural to human experience, so its interpretation (X) is only a noisy indicator of the underlying hedonic attribute levels ($\mathbf{x}^*$) even when no information is lost

[2] For example, the brightness level $\mathbf{x}^*$ of a light bulb is one of its hedonic attributes. Approximations or relevant measures for $\mathbf{x}^*$ are printed on its label in measurement units, say in this case, of 50 lumens. The observed signal X is then the agent's out-of-context interpretation or mental visualization of 50 lumens of brightness.

in the measurement step.[3] A person may have different interpretations of the same label information in different environments at different times. Therefore, although the print on the label is fixed, the readings or interpretations (X) made by an agent can be modeled as random.

Our friction assumes that conditional on the hedonic attributes, the noisy signals across different alternatives are correlated. This causes the agent to make different inferences about $\mathbf{x}^*$ when she faces different alternatives. To be more specific, the agent's information depends on context exogenously. In the choice problem $\{\mathbf{x}, \mathbf{y}\}$, the agent observes the signals X, Y, and in the choice problem $\{\mathbf{x}, \mathbf{z}\}$, the agent observes X, Z. Such dependency alone does *not* imply contextual choices.[4] The important assumption is that $X|\mathbf{x}^*$, $Y|\mathbf{y}^*$, and $Z|\mathbf{z}^*$ are correlated. When the agent observes X, Y in the choice problem $\{\mathbf{x}, \mathbf{y}\}$, she forms a posterior belief, say about $\mathbf{x}^*$, conditional on both signals X, Y. Because of the correlation, Y can provide additional information about $\mathbf{x}^*$, and so $\mathbf{x}^*|X, Y \neq \mathbf{x}^*|X$. When she faces the choice problem $\{\mathbf{x}, \mathbf{z}\}$, her posterior belief about $\mathbf{x}^*$ is conditioned on both X, Z. These two posterior beliefs, $\mathbf{x}^*|X, Y$ and $\mathbf{x}^*|X, Z$, are generally different when $\mathbf{y}^* \neq \mathbf{z}^*$, as are the posterior expected utilities of $\mathbf{x}$ in these two choice problems. Therefore, even if the agent is indifferent between $\{\mathbf{x}, \mathbf{y}\}$ and indifferent between $\{\mathbf{x}, \mathbf{z}\}$, she typically would *not* be indifferent between $\{\mathbf{y}, \mathbf{z}\}$. This simple observation can lead to intransitive choices.

The main novelty in our friction is that the correlation of the noisy signals is restricted to be positive across alternatives. We term this type of noise the *imperfect perception of attributes* as it is essentially specific to each attribute but not to each alternative. One may think of it as a misinterpretation of the relevant scientific units. If the noise biases an object's attribute upwards (downwards), it also biases upwards (downwards) perceptions of the same attribute of other objects. Given such noise, it becomes harder to perceive the absolute value of an attribute of an object than to perceive the relative difference between objects. Note that this noise in attributes is qualitatively different from noise in utilities. As shown in Proposition 3, our model does not satisfy monotonicity, and hence *cannot* be interpreted as a random utility model.

Such positively correlated noise can arise even when the label information is seemingly simple. Take choosing apartments as an example. Suppose an important hedonic attribute for the decision maker is the safety of the neighborhood. She can obtain a sense of safety by consulting the yearly crime statistics published by the local authority.[5] For each neighborhood, its safety can be measured in simple units such as the "number of crimes per year per ten thousand people" (the label information), but the decision maker's reading (X) of these numbers is still a noisy signal of safety ($\mathbf{x}^*$). For instance, it is not clear how strict the definition of crime is according to the authority. The decision maker's interpretation can be an exaggeration

[3] For example, different people would visualize different brightness levels (X) in their minds after reading "50 lumens".

[4] Indeed, when $X|\mathbf{x}^*$, $Y|\mathbf{y}^*$, and $Z|\mathbf{z}^*$ are independent, Y and Z do not provide any information about $\mathbf{x}^*$ and vice versa. Hence the posterior expected utilities of $\mathbf{x}$ in both $\{\mathbf{x}, \mathbf{y}\}$ and $\{\mathbf{x}, \mathbf{z}\}$ are equal in distribution. The same holds for both $\mathbf{y}$ and $\mathbf{z}$. The agent's choice reduces to a random utility model.

[5] For example, the local police department and city websites or the *Uniform Crime Reports* published by the FBI in the US.

(understatement) of safety for all neighborhoods if the local authority applies a narrower (broader) definition of crime than the decision maker realizes. Because the decision maker's perception is either exaggerated for all neighborhoods or understated for all neighborhoods, this uncertainty in interpretation results in the positively correlated imperfect perceptions. The decision maker can easily compare which neighborhood is safer, but she has more difficulty knowing the exact safety levels. Similar uncertainty in perceptions can occur for almost any attribute.[6] In general, imperfect perceptions arise whenever the decision-making agent *believes* that there is a common component in the uncertainty in her attribute perceptions.

The imperfect perception assumption is also supported by the *contrast effect* in psychology. The contrast effect is a well-known psychological phenomenon that refers to the enhancement or diminishment of the perception of any attribute when the object is contrasted with other objects weaker or stronger in the same attribute.[7] To illustrate the contrast effect in our model with the apartment example, suppose that the decision maker visited two apartments $\mathbf{x}$ and $\mathbf{y}$ with published neighborhood crime rates 5‰ and 8‰ (label information) respectively. Then, she researches a third apartment $\mathbf{z}$ and discovers that the crime rate is 1‰. It is much safer than both $\mathbf{x}$ and $\mathbf{y}$. Since 1‰ is unusually low, the decision maker infers that it is unlikely that the local authority adopts a very broad definition of crime. Therefore, a narrower definition (than the decision maker's) is probably being used, causing a downward bias in the agent's interpretations (X, Y, and Z) of the data. Hence, after checking the data for $\mathbf{z}$, she revises her perception of the safety of both $\mathbf{x}$ and $\mathbf{y}$, and they are perceived as less safe than before. This is in accordance with the contrast effect, which occurs when a very safe option $\mathbf{z}$ is available, causing $\mathbf{x}$ and $\mathbf{y}$ to be perceived as less safe. In Section 2, we elaborate further that the existence of positively correlated imperfect perceptions is also supported by the random anchoring phenomenon found in psychology.

When the decision maker's preferences are monotonically determined by a single attribute, the contrast effect is inconsequential: she always chooses to maximize (or minimize) that attribute.[8] However, if her preference involves at least two attributes, a third competing option $\mathbf{z}$ can affect her perception of the two attributes differently and simultaneously increase one and decrease the other. Hence, the relative utilities of $\mathbf{x}$ and $\mathbf{y}$ can change when contrasted with $\mathbf{z}$.

One example is the compromise effect, and below we explain intuitively how our model addresses it. Suppose in choosing apartments, the decision maker faces a trade-off between neighborhood safety and

[6] The noise can also arise naturally in the perception of nonnumerical information. Imagine that in addition to safety, the decision maker also prefers apartments with abundant natural light. She visits two apartments on the same day and sees that apartment $\mathbf{x}$ is brighter than $\mathbf{y}$ (i.e. X is brighter than Y). Although she does not know how bright the apartments typically are (she does not observe $\mathbf{x}^*, \mathbf{y}^*$), she learns the noisy signals (X and Y) from her visits. Each signal may be inaccurate, but the difference between signals can clearly indicate which room is typically brighter. After all, she is seeing both apartments at roughly the same time, in the same weather. There is a natural common component in the noise of the signals. The same intuition holds in perceiving other attributes, such as the noisiness of the neighborhood and the length of commuting time.

[7] See e.g. Schwarz and Bless (1992) and Plous (1993), pages 38 - 41.

[8] That is, if the agent only cares about safety, she always chooses the safest apartment.

energy efficiency. She prefers (in hedonic attributes) higher safety as well as higher efficiency. There are two apartments $\mathbf{x}$ and $\mathbf{y}$ and our decision maker reads the relevant label information about the two attributes. Suppose $\mathbf{x}$ is in a safer neighborhood (crime rate of 5‰) but its measured efficiency level is medium, whereas $\mathbf{y}$ is slightly less safe (crime rate of 8‰) but has high efficiency. Suppose that between the two, the decision maker is inclined to choose $\mathbf{y}$. Now introduce a third option $\mathbf{z}$ that is in a very safe neighborhood (crime rate of 1‰), but its energy efficiency is measured to be very low. According to the compromise effect, introducing $\mathbf{z}$ makes $\mathbf{x}$ a compromise between $\mathbf{z}$ and $\mathbf{y}$, and $\mathbf{x}$ is chosen more often. In our model, due to the contrast effect, the (posterior) perceived safety of both $\mathbf{x}$ and $\mathbf{y}$ is reduced after $\mathbf{z}$ is introduced. Again, due to the contrast effect, the (posterior) perceived energy efficiency of both $\mathbf{x}$ and $\mathbf{y}$ increases after $\mathbf{z}$ is introduced. Now, reducing the perceived safety of $\mathbf{x}$ and $\mathbf{y}$ affects both apartments negatively, but more so for $\mathbf{y}$ because of diminishing marginal utility in safety. Increasing the perceived efficiency of $\mathbf{x}$ and $\mathbf{y}$ affects both apartments positively, but more so for $\mathbf{x}$, due to the diminishing marginal utility in energy efficiency. Consequently, $\mathbf{x}$ has a higher expected utility level relative to $\mathbf{y}$ after $\mathbf{z}$ is introduced, and is chosen more often.

In addition to the assumption of imperfect perceptions, Bayesian updating is also an important component of our model. We adopt Bayesian updating because it is the canonical benchmark in modeling information and learning. If there is no updating at all, presenting the alternative $\mathbf{z}$ will not affect perceptions of $\mathbf{x}$ and $\mathbf{y}$. Despite its importance, we do not claim that people in reality perform sophisticated Bayesian updating and calculate posterior expectations. Instead, we interpret the model as an as-if representation of the decision process. As seen in the example above, this as-if process does parallel an intuitive explanation of contextual choices.

We present the general set-up in Section 2. In Section 3, we apply a special parametric case of the model to explain contextual choices in detail. An analysis of the general model is presented in Section 4, where we study the decoy choice pattern, choices related to dominating options, and some extensions of the model. Section 5 contains further discussion and the limitations of the model. Additional proofs are in the appendix.

1.1. Related Literature

A number of contextual effects can be explained by existing models in the literature using context-independent preferences. Our paper contributes to this literature by proposing a new and parsimonious informational mechanism that complements existing explanations. In particular, our model explicitly connects contextual choices with the underlying attribute space. Therefore, our model not only explains these contextual effects but also explains why these effects only happen for certain options in the attribute space. The following briefly reviews the contextual effects we focus on and some of the related context-independent utility models in the literature.

- Intransitivity (Tversky 1969) can be accommodated by random utility models such as the mixed logit model via the Condorcet cycle. Some other models whose choice probabilities contain a random utility component, such as the model in Guo (2016), can also accommodate intransitivity. Other notable models for intransitivity can be found in Table 5 of Rieskamp et al. (2006). We describe intransitive choices in detail and illustrate how our model addresses this phenomenon in Section 3.2.
- J-s reversal (Hsee 1996) can be accommodated by the models in Wernerfelt (1995) and Kamenica (2008), but not by random utility models. Since j-s reversal is an evaluation problem instead of a choice problem, many models on choice do not directly address this phenomenon. We describe j-s reversal in detail and illustrate how our model addresses this phenomenon in Section 3.3.
- The compromise effect (Simonson 1989) can be accommodated by several models, including those of Wernerfelt (1995), Kamenica (2008), Guo (2016), and Natenzon (2019), but cannot be addressed by random utility models because of the monotonicity property. The models that accommodate the compromise effect usually also accommodate the attraction effect of Huber et al. (1982). However, very few models consider unchoosable alternatives, so the phantom decoy effect (Pratkanis and Farquhar 1992) is unaddressed. We describe and address the compromise effect in detail and its relation to the monotonicity property in Section 3.4. In Section 4.1, we define a comparative static for the decoy choice pattern that unifies the compromise effect, the attraction effect, and the phantom decoy effect. We show in Theorem 1 that our model satisfies this decoy choice pattern.
- We also briefly discuss how our model addresses the similarity effect (Tversky and Russo 1969) in Section 4.4. This effect can be accomodated by most of the papers in this literature.

Our paper is related to Wernerfelt (1995), Kamenica (2008), Guo (2016), and Natenzon (2019). Guo (2016) and Natenzon (2019) both study single agent decisions under informational frictions. Their information structures are not specific to attributes and hence are different from ours. The friction in Guo (2016) is the uncertainty about individual preferences. The model endogenizes the effort to learn one's preference in different contexts. Different from Guo (2016), we do not endogenize information acquisition; hence our mechanism is not related to phenomena in which the selective acquisition of information is the main concern.[9] Natenzon (2019) studies a transitive choice model in which the agent directly receives a signal about the utility levels of each option. In contrast, our model studies a different information structure and can explain intransitive choices. One important distinction of our approach from the literature is that we directly model attribute information. Therefore our model naturally identifies (through attributes) the types of choice problems in which contextual effects can (or cannot) arise. For example, our model complements the above theories by predicting that these contextual effects arise only when there are at least two attribute dimensions under consideration. Our results also state various contextual choices in terms of comparative

[9] For example, see Guo (2016)'s explanation of choice overload.

statics in both attributes and choice probabilities, instead of choice probabilities alone. In this sense, our paper is closest to Wernerfelt (1995) and Kamenica (2008) who also model information from more than one attribute. However, in their paper learning is the equilibrium outcome of a consumer-retailer market game. In contrast, our model does not require strategic considerations for contextual choices to occur.

Our model also differs from the class of random utility models, which includes Block and Marschak (1960), Falmagne (1978), Thurstone (1927), Luce (1959), Tversky (1972), Hausman and Wise (1978) and more recently, Gul et al. (2014). As detailed in Section 3.3, because random utility models are monotonic, they cannot explain the increase in absolute choice probabilities in the compromise effect.

Our model is also related to another stream of the literature on reference-dependent preferences in which utilities are directly assumed to depend on the choice set. See e.g. Simonson (1989), Tversky and Simonson (1993), Koszegi and Rabin (2006), Bordalo et al. (2013), Ok et al. (2015) and Tserenjigmid (2019). This is a broad class of models that assumes a context-dependent utility function with a reference point. Many of these models, however, do not explicitly model how the reference point changes from one choice problem to another. When the reference point is left as a free parameter, a reference-dependent model typically has high explanatory power but lacks predictive power — it does not provide testable implications. On the other hand, their prediction power can be defined and tested in empirical studies in which an obvious reference option is given by design. Our paper complements such models in studying decision problems that involve no obvious reference point by design.

We now focus our comparison with the class of reference point models in which the reference point is explicitly modeled. To the best of our knowledge, the most notable examples are Koszegi and Rabin (2006), Bordalo et al. (2013), Ok et al. (2015) and Tserenjigmid (2019). Our model is qualitatively different from all these models. The model of Koszegi and Rabin (2006) does not capture the compromise effect, and hence it is qualitatively different from our model. The model of Bordalo et al. (2013) is the most general in the sense that it allows for both the compromise effect and the reverse of the compromise effect, as well as both the attraction effect and its reverse. Such generality distinguishes their model from ours. The model in Ok et al. (2015) does not allow pairwise cycles, a prediction that is not ruled out by our model. The model of Tserenjigmid (2019) also differs from ours since it does not directly explain the j-s reversal. We explain the above differences in detail in Section 4.6.

Apart from comparing with these existing models, we ask a further question that in terms of qualitative predictions, how does our model relate to this whole class of reference point models. We find that in terms predicting choices, there is a non-empty intersection. In Section 4.5, we derive a reference point model (outside existing literature) that is choice-equivalent to a limiting case of our model. However, we show in the same section that this reference point model cannot explain evaluation tasks such as the j-s reversal. This relation between our model and the class of reference-dependent models is also a conceptual contribution of our paper: a very simple/minimal informational friction under standard Bayesian updating of beliefs can be shown to explain several contextual choices without directly assuming a reference point. Moreover, our approach enjoys the important advantage of a stable underlying utility function. Through our model, a fixed utility function can be estimated from the choice data. Taking this utility function as a behavioral benchmark, welfare analysis can then be easily carried out.

2. The Model, Its Assumptions and Its Motivations

Empirically, contextual choices occur only when options have *two or more* attributes. Therefore, we take the primitives of our model to be the hedonic attributes of each object (which are not precisely observed by the agent) and use $\mathbb{R}^n$ for $n \geq 2$ to represent the attribute space. The attributes of each item $\mathbf{x}$ are represented as a vector $\mathbf{x}^* := (x_1^*, \ldots, x_n^*)$ in that space, with each coordinate given by the corresponding attribute level. The restriction to $\mathbb{R}^2$ is for expositional simplicity, and the proofs of the main results can be extended to higher dimensions in a straightforward manner.

The vector of attributes $\mathbf{x}^*$ is not directly observed by the agent. The agent observes noisy signals of the attributes and tries to maximize her payoff given those signals. We assume that the agent has context-independent preferences over the attribute space that can be represented by a vNM utility function. Following standard consumer theory, we assume that her preferences are monotonic in all attributes. Additionally, utility has diminishing returns, and there is weak complementarity between attributes. We call a preference *standard* if it displays these properties.

ASSUMPTION 1 (**STANDARD PREFERENCE**). *The decision maker's preference over the distributions on* $\mathbb{R}^2$ *can be represented by a vNM utility function* $u : \mathbb{R}^2 \to \mathbb{R}$ *that is differentiable, increasing (i.e.,* $u_1 > 0, u_2 > 0$*), and exhibits decreasing marginal sensitivity (i.e.,* $u_{11} < 0, u_{22} < 0$*) and weak complementarity (i.e.,* $u_{12} \geq 0$ *). Any utility function representing a standard preference is called a standard utility function.*

Moreover, the agent is Bayesian with a prior belief over $\mathbb{R}^2$. The prior distribution represents the agent's anticipation of attribute levels before she observes any signals. We endow the agent with a normal prior distribution. Without loss of generality, we translate and scale the attribute space such that the prior mean is at the origin and the prior variance is $\Omega := \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$ for some $r \in (-1, 1)$.[10] After receiving the signals, the agent chooses an option to maximize her posterior expected utility.

ASSUMPTION 2 (**NORMAL-BAYESIAN AGENT**). *The decision maker is Bayesian with a normal prior* $\mathcal{N}(0, \Omega)$ *and maximizes her posterior expected utility.*

Next, we assume a novel type of noise in the perception of attributes. The noise is specific only to the attributes, and hence is common across alternatives. Let capital letters (i.e. $X = (X_1, X_2)$) denote the noisy signal of the attributes $\mathbf{x}^*$. For instance, in the choice set $\{\mathbf{x}, \mathbf{y}\}$, the attribute levels $\mathbf{x}^*$ and $\mathbf{y}^*$ are

[10] Such a correlation can arise when, for example, the two attributes are price and quality. One can interpret $r < 0$ as the agent having a prior belief that a good price is associated with low quality.

signaled by $X = \mathbf{x}^* + \epsilon$ and $Y = \mathbf{y}^* + \epsilon$, where ϵ has the same realization for all objects. Hence the agent perceives the relative differences in attributes between the items, i.e., $\mathbf{x}^* - \mathbf{y}^* = X - Y$, better than the absolute locations $\mathbf{x}^*$ and $\mathbf{y}^*$ in the attribute space. This common noise across alternatives is assumed for mathematical simplicity and can be relaxed. Section 4.3 shows that the noise does not have to be identical across alternatives. Instead, the noise for each attribute needs to be positively correlated across alternatives. In addition, although the noise is additive, our model can also cover the case of multiplicative noise by logarithmically transforming the attributes.[11]

Assumption 3 (**Imperfect Perception**). *For any n alternatives $\{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ each with attributes $\mathbf{x}^{1*}, \ldots, \mathbf{x}^{n*} \in \mathbb{R}^2$, the agent receives signals $X^1, \ldots, X^n$, where $X^i - \mathbf{x}^{i*} = \epsilon$ for all i. The noise term $\epsilon \sim \mathcal{N}(0, T^{-1})$ is normal with variance matrix*

$$T^{-1} = \begin{bmatrix} 1/t_1^2 & R/(t_1 t_2) \\ R/(t_1 t_2) & 1/t_2^2 \end{bmatrix} \textit{ for some } \frac{1}{t_1^2} + \frac{1}{t_2^2} > 0, \textit{ and some } R \in (-1, 1).$$

Several motivating arguments can be made for this assumption. As discussed in the introduction, our perception of attributes is susceptible to imperfections even when the attributes are measured and described in numbers and text. Here, $\mathbf{x}^*$ represents a vector of hedonic attributes that enter the utility function. It can include different sensory information, such as brightness or apparent temperature. Some qualities or the intensity of $\mathbf{x}^*$ can be approximated or measured in units such as lumen or degrees Celsius. When people read and interpret these measurements, they obtain a signal X for $\mathbf{x}^*$ by interpreting these numbers back into sensory information. Experiments show that the people are *not* able to interpret these numerical information perfectly (Green and Srinivasan 1978, Ariely et al. 2003). For instance, Ariely et al. (2003) find that simply hearing a sound provides a better perception of the volume than reading the measured volume in numbers. Pictorial descriptions of an item generate as precise, if not more precise perceptions of its attributes than numerical and verbal descriptions do (Green and Srinivasan 1978). One explanation for these findings is that the decision maker is subjectively uncertain in interpreting units. As Kamenica (2008) argues, in general *"interpreting technical units of quality can be difficult."* For instance, a person who is used to seeing temperature in degrees Celsius finds it hard to interpret degrees Fahrenheit. In fact, even in degrees Celsius, the same person's interpretation of numeric temperatures is noisy. Due to such difficulty, reading precise measurements provides only a noisy indicator of the hedonic attribute levels. When the decision maker is uncertain in interpreting technical units, her interpretation of the measured attributes may be lower or higher than their true level, resulting in an under or overperceived attribute across all alternatives. In this way, a decision maker would believe there is a common component in her noisy perceptions across all options.

[11] More precisely, suppose the noisy signal is defined by $X = \mathbf{x}^* \times e^{\epsilon}$, where ϵ is the same across alternatives. This multiplicative specification of the signal maintains the attribute ratio instead of the difference between alternatives. In this case, one can first apply a log-transformation so that $\ln X = \ln \mathbf{x}^* + \epsilon$. Our model is then directly applicable to the transformed variables.

There is also an instrumental argument for imperfect perception: people's perceptions are inconsistent in the same way as if they were affected by imperfect perceptions. Indeed, some experiments in Ariely et al. (2003) show that people's perceptions can be affected by a "random anchor", causing "coherent arbitrariness". In evaluation tasks, participants usually evaluate the absolute value of an attribute level arbitrarily, but differences in evaluations across alternatives is coherent with the differences in their attribute levels. This finding is robust to whether the attributes are displayed in technical units or not. As stated by Ariely et al. (2003),

> *"[W]e show that consumers' absolute valuation of experience goods is surprisingly arbitrary, even under 'full information' conditions. However, we also show that consumers' relative valuations of different amounts of the good appear orderly . . . "*[12]

In light of these findings, our imperfect perception assumption can be interpreted as a random anchor affecting the perception of attributes. Hence, the agent either overperceives or underperceives each attribute across all alternatives.

There may be other arguments for what causes this particular noise in perception. Nevertheless, the explanation in our model holds *as long as the decision maker believes* there can be a common noise in her perception of the attributes. Regardless of whether her belief is true, contextual effects can occur through posterior utility maximization when the agent forms her posterior taking into consideration the imperfect perception.

Normality is adopted for prior-signal conjugacy. We allow the standard deviations to differ across attributes as long as one of them is strictly positive (i.e. $\frac{1}{t_1^2} + \frac{1}{t_2^2} > 0$), and the other can be zero (e.g. $t_1 = \infty$). Our assumption also allows the noise across attributes to be correlated (R).[13]

Below summarizes some of the notations. Bold letters (e.g., $\mathbf{x}, \mathbf{y}, \mathbf{z}$) denote different alternatives. Letters with an asterisk (e.g., $\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*$) denote the (unobserved) hedonic attribute levels of objects in $\mathbb{R}^2$. Capital letters (e.g., X, Y, Z) denote their respective signals. We denote more than three alternatives with superscripts. Calligraphic letters (i.e., $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$) denote the agent's posterior beliefs about the hedonic attributes. Subscripts distinguish the respective attribute dimensions for a given vector. We use $C(\mathbf{x}^l, \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^i, (\mathbf{x}^{i+1}, \ldots, \mathbf{x}^{i+j})\})$ to denote the choice probability of $\mathbf{x}^l$ from the set $\{\mathbf{x}^1 \ldots \mathbf{x}^{i+j}\}$ in which $\{\mathbf{x}^{i+1}, \ldots, \mathbf{x}^{i+j}\}$ are unavailable. A $C(.,.)$ that assigns a probability to any $\mathbf{x}$ in every nonempty

[12] One can interpret these observations as follows. When an attribute level of an option is perceived to be higher (and hence of higher value), the attribute level for other options is also perceived to be higher (and so are also of higher value). However, the differences in the perceived levels (and in the values) is consistent with the differences in the actual attribute levels among the options. Hence, differences between options are perceived coherently, but the absolute value of the attribute is perceived arbitrarily.

[13] Such a correlation can arise when attributes are closely related, such as the sugar content and calories in a soft drink, and one might expect such a correlation in the noise across these attributes.

finite set of alternatives S, with any $S' \subsetneq S$ specifying the unavailable objects, is called the *choice behavior* of an agent. The choice behavior satisfies

$$\sum_{k=1}^{i} C(\mathbf{x}^k, \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^i, (\mathbf{x}^{i+1}, \ldots, \mathbf{x}^{i+j})\}) = 1.$$

3. A Special Parametric Case

In this section, we illustrate stochastic intransitivity, j-s reversal and the compromise effect with the following parametric setting. The stable preference is described by the simple exponential utility $u : \mathbb{R}^2 \to \mathbb{R}$

$$u(x_1, x_2) = -e^{-3x_1} - e^{-3x_2}.$$

We adopt this simple CARA utility function in this section because it works well with the normal distribution and provides an analytical expression for the posterior utility and choice probabilities. In this section, the noise structure is simple and one dimensional. The first attribute is perfectly interpreted and perceived without noise. Noise exists only in the second attribute. Mathematically, this is expressed as

$$\epsilon \sim \mathcal{N}\left(0, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right).$$

Finally, the agent's prior is taken to follow the standard bivariate normal distribution centered at the origin.

3.1. The Contrast Effect

Before addressing the choice effects, we first illustrate how the contrast effect is manifested in the model. In this parametric setting, the first attribute is perceived noiselessly so the contrast effect only occurs for the second attribute. Since $X - \mathbf{x}^* = Y - \mathbf{y}^*$, a straightforward calculation of the posterior belief from Bayesian updating gives

$$\mathcal{X}_1 | X, Y = x_1^*, \text{ and } \mathcal{X}_2 | X, Y \sim \mathcal{N}\left(\frac{1}{3}(2X_2 - Y_2), \frac{1}{3}\right).$$ [14]

The agent's belief about the first attribute, $\mathcal{X}_1 | X, Y$, exactly equals its hedonic attribute level x_1^* since it is perceived noiselessly. Her belief about the second attribute exhibits the contrast effect. Under the prior, the probability of $\mathbf{y}_2^*$ being extremely high is limited. If $\mathbf{y}$ is surprisingly high in the second attribute (i.e., then Y_2 is also high), the posterior belief then assigns larger probability on ϵ_2 being high (so that it makes sense to see a high Y_2 when $\mathbf{y}_2^*$ is low in the prior). Because $X_2 - \mathbf{x}_2^* = Y_2 - \mathbf{y}_2^* = \epsilon_2$, this creates the contrast effect in perception: $\mathbf{x}$ is perceived to be poorer in the second attribute. After the introduction of $\mathbf{y}$, the agent thinks that ϵ_2 is high. Her posterior for $\mathbf{x}_2^*$ under the same X_2 needs to be adjusted downwards because $\mathbf{x}_2^* = X_2 - \epsilon_2$. The better $\mathbf{y}$ is in the second attribute, the worse $\mathbf{x}$ is in the second attribute in the agent's posterior belief.

[14] Similarly, $\mathcal{Y}_1 | X, Y = \mathbf{y}_1^*$, and $\mathcal{Y}_2 | X, Y \sim \mathcal{N}\left(\frac{1}{3}(2Y_2 - X_2), \frac{1}{3}\right)$.

3.2. Violation of Weak Stochastic Transitivity

Weak stochastic transitivity refers to the postulate that if $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) \geq 0.5$ and $C(\mathbf{y},\{\mathbf{y},\mathbf{z}\}) \geq 0.5$, then $C(\mathbf{x},\{\mathbf{x},\mathbf{z}\}) \geq 0.5$. Early evidence of its violation can be found in Tversky (1969), and more recent evidence in Rieskamp et al. (2006). They find that weak transitivity can be violated when there is *no clear dominance among* $\mathbf{x},\mathbf{y},\mathbf{z}$ *in attributes*, which we illustrate in the proposition below.

In our model, intransitivity results from the crossing of stochastic indifference curves. Due to the randomness ϵ in the information, the choice between any two objects $\mathbf{x}$ and $\mathbf{y}$ depends on their attribute levels $\mathbf{x}^*,\mathbf{y}^*$ and the realization of ϵ. Hence, given their attribute levels, we can determine the probability of choice, $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\})$, from the distribution of ϵ. We say that the agent is *stochastically indifferent* between $\mathbf{x}$ and $\mathbf{y}$ (written $\mathbf{x}\sim\mathbf{y}$) if

$$C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) = 0.5.$$

Similarly, the *stochastic indifference curve* for $\mathbf{x}$ is the set of alternatives to which the agent is stochastically indifferent relative to $\mathbf{x}$. In the space of attributes, this set of alternatives corresponds to the following set of attributes $\{\mathbf{y}^* \in \mathbb{R}^2 | \mathbf{x}\sim\mathbf{y}\}$.

Consider two alternatives $\mathbf{x},\mathbf{y}$ such that $x_1^* > y_1^*$ and $y_2^* > x_2^*$. When is $\mathbf{x}$ chosen over $\mathbf{y}$? Since the agent is Bayesian, she chooses $\mathbf{x}$ whenever the posterior expected utility of $\mathbf{x}$ is greater than that of $\mathbf{y}$. In our notation, the agent's posterior beliefs about $\mathbf{x}^*$ and $\mathbf{y}^*$ are the random variables $\mathcal{X}|X,Y$ and $\mathcal{Y}|X,Y$ respectively. So $\mathbf{x}$ is chosen over $\mathbf{y}$ if and only if

$$\mathbb{E}[u(\mathcal{X})|X,Y] > \mathbb{E}[u(\mathcal{Y})|X,Y].$$

Substituting the posterior into the expected utility formula gives that $\mathbf{x}$ is chosen over $\mathbf{y}$ if and only if

$$\mathbb{E}[u(\mathcal{X})|X,Y] = -e^{-3x_1^*} - e^{-(2X_2-Y_2)+3/2} > -e^{-3y_1^*} - e^{-(2Y_2-X_2)+3/2} = \mathbb{E}[u(\mathcal{Y})|X,Y].$$ [15]

To obtain the choice probability, substitute $X-\mathbf{x}^* = Y-\mathbf{y}^* = \epsilon$ to obtain the equivalent inequality

$$-\frac{3}{2} + \ln\left(\frac{e^{-3y_1^*} - e^{-3x_1^*}}{e^{y_2^*-2x_2^*} - e^{x_2^*-2y_2^*}}\right) > -\epsilon_2.$$

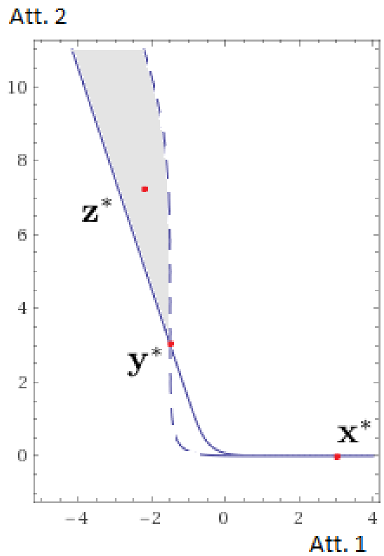
Since $\epsilon_2 \sim \mathcal{N}(0,1)$, the choice probability can be expressed using the normal c.d.f Φ,

$$C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) = \Phi\left(-\frac{3}{2} + \ln\left(\frac{e^{-3y_1^*} - e^{-3x_1^*}}{e^{y_2^*-2x_2^*} - e^{x_2^*-2y_2^*}}\right)\right).$$

For interpretation, first recall that $x_1^* > y_1^*$ and $y_2^* > x_2^*$. Therefore, both $e^{-3y_1^*} - e^{-3x_1^*}$ and $e^{y_2^*-2x_2^*} - e^{x_2^*-2y_2^*}$ are positive. Moreover, since both Φ and $\ln$ are increasing functions, the choice probability is increasing in x_1^* and x_2^* and decreasing in y_1^* and y_2^*. Intuitively, the agent is more likely to choose $\mathbf{x}$ if the hedonic attribute levels of $\mathbf{x}$ improve and is less so if the hedonic attributes of $\mathbf{y}$ become more desirable.[16]

[15] The expected utilities above can be understood as follow. In $\mathbb{E}[u(\mathcal{X})|X,Y] = -e^{-3x_1^*} - e^{-(2X_2-Y_2)+3/2}$, the utility from the first attribute is clear due to perfect perception. We have mentioned that the contrast effect influences perception in the second attribute, and hence also the expected utility. The better Y_2 is, the smaller the expected utility for $\mathbf{x}$. The constant in the exponent of the second term comes from the uncertainty. Because $\mathcal{X}_2|X,Y$ is normally distributed, $e^{-3\mathcal{X}_2}$ is log-normal, and its expectation involves a constant from the variance of $\mathcal{X}_2|X,Y$.

[16] We show that if $x_1^* > y_1^*$ and $y_2^* > x_2^*$ do not hold, the dominating option is chosen with probability 1 in the next section.



Empirically, intransitivity can happen when the difference is easier to discriminate in attribute one, and harder in attribute two (Tversky 1969, Leland 1994). I.e. if the difference in attribute two is not large enough to be "consequential", individuals choose the option higher in attribute one. However, individuals choose the option higher in attribute two if its difference is large enough.

In our model, such observation can happen locally near $\mathbf{y}$ *as shown by the indifference curves of* $\mathbf{y}$ *(dashed) and of* $\mathbf{x}$ *(solid). For options slightly worse than* $\mathbf{y}$ *in attribute one, a large difference in attribute two is needed for them to compare favourably to* $\mathbf{y}$*. And for options (slightly) better than* $\mathbf{y}$ *in attribute one, also a sizable difference in attribute two of approximately* $y_2^* - x_2^*$ *is needed to compare unfavourably to* $\mathbf{y}$*. Now that* $z_2^* - x_2^* > y_2^* - x_2^*$ *is more than enough,* $\mathbf{x}$ *compares unfavourably to* $\mathbf{z}$ *even though* x_1^* *is mucb better than* z_1^**.*

Figure 1 **Crossing Stochastic Indifference Curves**

The indifference curve can be traced out using the definition $C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\}) = 0.5$. Because $\Phi(0) = 0.5$, we have $\mathbf{x} \sim \mathbf{y}$ if and only if

$$0 = -\frac{3}{2} + \ln\left(\frac{e^{-3y_1^*} - e^{-3x_1^*}}{e^{y_2^* - 2x_2^*} - e^{x_2^* - 2y_2^*}}\right).$$

The agent is stochastically indifferent between any $\mathbf{x}$ and $\mathbf{y}$ with attributes satisfying the above equation.

Generically, if $\mathbf{x} \sim \mathbf{y}$, their indifference curves cross. For illustration, we let $\mathbf{x}^* = (3, 0)$ and $\mathbf{y}^* = (3 - \frac{1}{3}\ln(1 - e^{9/2} + e^{27/2}), 3)$ and check that $\mathbf{x} \sim \mathbf{y}$. As shown in Figure 1, the red dots are the corresponding hedonic attribute levels, and the indifference curve of $\mathbf{x}$ is the solid curve, whereas that of $\mathbf{y}$ is dashed. The two curves intersect at $\mathbf{x}^*$ and $\mathbf{y}^*$. The curves are indistinguishable for large values in the first attribute. Because the curves are distinct, intransitivity can occur when we consider any $\mathbf{z}$ with attributes in the shaded area. As in Figure 1, $\mathbf{z}^*$ is below the $\mathbf{y}$-curve and above the $\mathbf{x}$-curve. So $C(\mathbf{y}, \{\mathbf{y}, \mathbf{z}\}) > 0.5$ and $C(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) < 0.5$. However, as can be easily seen, slightly improving $\mathbf{x}^*$ in either attribute will cause $C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\}) > 0.5$, thereby strictly violating weak transitivity. The above analysis is a proof of the following existence result.

Proposition 1. *Suppose there is imperfect perception of one of the attributes. There is a normal-Bayesian agent with a standard preference that is intransitive over* $\mathbf{x}$, $\mathbf{y}$, *and* $\mathbf{z}$ *for some* $x_1^* > y_1^* > z_1^*$ *and* $z_2^* > y_2^* > x_2^*$.

In Section 4, we show that intransitivity cannot occur when there is a dominance relationship between the alternatives in terms of attributes.

3.3. Joint-Separate Evaluation Reversal

This effect refers to the reversal of the evaluations of alternatives in two contexts. In an experiment of Hsee (1996), the subjects (as company owners) were asked for valuations in terms of willingness to pay to hire different job candidates as programmers. Candidate $\mathbf{x}$ had a college GPA of 4.9 out of 5 and had written 10 programs in the computer language KY. Candidate $\mathbf{y}$ had a GPA of 3.0 from the same school, and had written 70 programs in the same language. When the subjects were asked to evaluate $\mathbf{x}$ alone, the average valuation was 32.7k dollars in salary; when asked to evaluate $\mathbf{y}$ alone, the average valuation was 26.8k. However, when the two candidates were presented together, the evaluations reversed. The average valuation for $\mathbf{x}$ in the presence of $\mathbf{y}$ became 31.2k, less than the new valuation of 33.2k for $\mathbf{y}$. With an abuse of notation, we denote by $\$(\mathbf{x})$ and $\$(\mathbf{y})$ the valuation or the average willingness to pay for $\mathbf{x}$ and $\mathbf{y}$ in dollars, and denote by $\$(\mathbf{x}|\mathbf{x},\mathbf{y})$ the average valuation for $\mathbf{x}$ in the presence of $\mathbf{y}$, and $\$(\mathbf{y}|\mathbf{x},\mathbf{y})$ for $\mathbf{y}$ in the presence of $\mathbf{x}$. A decision maker is said to *display j-s reversal* if there exist $\mathbf{x},\mathbf{y}$ such that both $\$(\mathbf{x}) > \$(\mathbf{y})$ and $\$(\mathbf{x}|\mathbf{x},\mathbf{y}) < \$(\mathbf{y}|\mathbf{x},\mathbf{y})$ holds.

In the experiment, the two attributes are GPA and programming experience. While GPA (scaled out of 5) is easy to interpret, programming experience is hard. Although programming experience is explicitly measured in the number of programs written, it is not clear how advanced the computer language KY is or how difficult it is to write programs in it. The subjects as “company owners” may not have been experts in programming, and may have been uncertain of their subjective interpretations. Hence, it is reasonable to model programming experience with imperfect perception.

To show the existence of reversal in our model, we need to find a pair of $\mathbf{x}$ and $\mathbf{y}$ such that $x_1^* > y_1^*$ and $x_2^* < y_2^*$, and that $\$(\mathbf{x}) > \$(\mathbf{y})$ and $\$(\mathbf{x}|\mathbf{x},\mathbf{y}) < \$(\mathbf{y}|\mathbf{x},\mathbf{y})$ hold simultaneously. In this subsection, we use the *average posterior expected utility* as a proxy for average willingness to pay. That is, $\$(\mathbf{x})$ is understood as the average posterior expected utility of $\mathbf{x}$ in $\{\mathbf{x}\}$, $\$(\mathbf{y})$ is that of $\mathbf{y}$ in $\{\mathbf{y}\}$, and $\$(\mathbf{x}|\mathbf{x},\mathbf{y})$, $\$(\mathbf{y}|\mathbf{x},\mathbf{y})$ that of $\mathbf{x}$ and of $\mathbf{y}$ in $\{\mathbf{x},\mathbf{y}\}$ respectively.

When there is only one option, the posterior is based only on its own signal. For noiseless perception, $\mathcal{X}_1|X = x_1^*$. The noisy perception has the Bayesian posterior $\mathcal{X}_2|X \sim \mathcal{N}(\frac{1}{2}X_2, \frac{1}{2})$. Hence the average posterior expected utility is

$$\$(\mathbf{x}) := \mathbb{E}_X[\mathbb{E}_{\mathcal{X}_2}[-e^{-3x_1^*} - e^{-3\mathcal{X}_2}|X]] = -e^{-3x_1^*} - e^{-\frac{3}{2}x_2^* + \frac{27}{8}}.$$ [17]

From a similar analysis in the previous subsection, when there are two options, we obtain

$$\$(\mathbf{x}|\mathbf{x},\mathbf{y}) := \mathbb{E}_{X,Y}[\mathbb{E}_{\mathcal{X}_2}[-e^{-3x_1^*} - e^{-3\mathcal{X}_2}|X,Y]] = -e^{-3x_1^*} - e^{-(2x_2^* - y_2^*)+2}.$$

Additionally, an analogous expression holds for $\$(\mathbf{y}|\mathbf{x},\mathbf{y})$. The two inequalities $\$(\mathbf{x}) > \$(\mathbf{y})$ and $\$(\mathbf{x}|\mathbf{x},\mathbf{y}) < \$(\mathbf{y}|\mathbf{x},\mathbf{y})$ are then

$$\begin{cases} -e^{-3x_1^*} - e^{-\frac{3}{2}x_2^* + \frac{27}{8}} > -e^{-3y_1^*} - e^{-\frac{3}{2}y_2^* + \frac{27}{8}} \\ -e^{-3x_1^*} - e^{-(2x_2^* - y_2^*)+2} < -e^{-3y_1^*} - e^{-(2y_2^* - x_2^*)+2}. \end{cases}$$

[17] A similar expression holds for $\mathbf{y}$.

There are many pairs of alternatives that satisfy both inequalities. For illustration, let $\mathbf{x}^*$ be $(3,0)$. Figure 2 plots the shaded region where both inequalities are satisfied. The dashed curve is the boundary defined by the first inequality above, and the solid curve is the boundary defined by the second. Any $\mathbf{y}$ with attributes $\mathbf{y}^*$ in the shaded region is an example of the desired reversal.

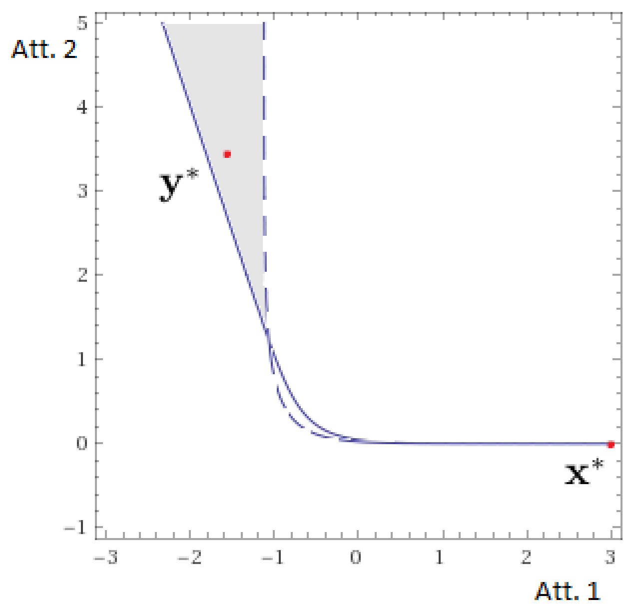


Figure 2 Joint-Separate Evaluation Reversal

Empirically, j-s reversal can be observed if attribute one is easier to evaluate than attribute two (Hsee et al. 1999). In separate evaluation, this attribute one primarily determines the evaluation outcome. In a joint evaluation, comparison allows better evaluation of the attribute two, increasing its impact on the evaluation outcome, leading to the reversal. In our model, this observation can be interpreted as the case where attribute one is perceived with less noise than attribute two. The figure indicates the equi-value curve of $\mathbf{x}$ *in separate evaluation (dashed), and that in joint evaluation (solid). When the difference* $x_1^* - y_1^*$ *is significant enough,* $\mathbf{x}$ *is easily better valued in separate evaluation, even for* $\mathbf{y}$ *with rather large* y_2^* *in the shaded area. However, in joint evaluation, attribute two is better valued and hence better substitutes for attribute one, as illustrated by the flatter solid curve. So a large enough* y_2^* *is enough to compensate for the difference in* $x_1^* - y_1^*$*, and overall* $\mathbf{y}$ *becomes better valued in joint evaluation (i.e. being above the solid cuve).*

The mechanism that causes this reversal is shown in Figure 2. A $\mathbf{y}$ that is low in the first attribute easily satisfies $\$(\mathbf{y}) < \$(\mathbf{x})$ in separate evaluations. Because the utility function is concave and perception is noisy, a strong y_2^* attribute cannot effectively increase the overall evaluation. However, in a joint evaluation, there is a clear contrast in the second attribute for which $x_2^* < y_2^*$. In comparison, $\mathcal{X}|X,Y$ is perceived as much worse, and $\mathcal{Y}|X,Y$ much better, resulting in the reversal. The previous calculation proves the following result.

Proposition 2. *Suppose there is imperfect perception of one of the attributes. There is a normal-Bayesian agent with a standard preference who displays j-s reversal over* $\mathbf{x}$ *and* $\mathbf{y}$ *for some* $x_1^* > y_1^*$ *and* $x_2^* < y_2^*$.

3.4. The Compromise Effect

The compromise effect involves choice problems over two and three options. As in Figure 3, suppose there is a binary choice problem with options $\mathbf{x}, \mathbf{y}$ where $\mathbf{x}$ is better than $\mathbf{y}$ in the first attribute but $\mathbf{y}$ is better in the second. The *compromise effect* (Simonson (1989)) refers to the effect of introducing a third option $\mathbf{z}$ in or near the region C where $\mathbf{z}^*$ is extremely favorable in the first attribute but extremely unfavorable in the second attribute. Empirically, at the introduction of $\mathbf{z}$, subjects are generally led to choose the "compromise option" $\mathbf{x}$, increasing its choice frequency. Mathematically, let the initial choice set be $\{\mathbf{x}, \mathbf{y}\}$ and the extended choice set be $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. We define the *compromise effect* to be $C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) > C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\})$ for some $z_1^* > x_1^* > y_1^*$ and $y_2^* > x_2^* > z_2^*$.

Let $\Pr$ denote the probability measure for ϵ. We have seen previously that

$$C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\}) = \Pr\left(\mathbb{E}[u(\mathcal{X})|X, Y] > \mathbb{E}[u(\mathcal{Y})|X, Y]\right) = \Pr\left(\epsilon_2 > \frac{3}{2} - \ln\left(\frac{e^{-3y_1^*} - e^{-3x_1^*}}{e^{y_2^* - 2x_2^*} - e^{x_2^* - 2y_2^*}}\right)\right), \quad (1)$$

Similarly, we can also express the ternary probability as

$$\begin{aligned} &C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) \\ &= \Pr\left(\Big\{\mathbb{E}[u(\mathcal{X})|X, Y, Z] > \mathbb{E}[u(\mathcal{Y})|X, Y, Z]\Big\} \cap \Big\{\mathbb{E}[u(\mathcal{X})|X, Y, Z] > \mathbb{E}[u(\mathcal{Z})|X, Y, Z]\Big\}\right), \end{aligned}$$

where the first term in the intersection is the event that $\mathbf{x}$ is perceived to be better than $\mathbf{y}$,

$$\{\mathbb{E}[u(\mathcal{X})|X, Y, Z] > \mathbb{E}[u(\mathcal{Y})|X, Y, Z]\} = \{\epsilon_2 > \frac{3}{2} - \frac{4}{3}\ln\left(\frac{(e^{-3y_1^*} - e^{-3x_1^*})}{e^{-\frac{3}{4}(3x_2^* - y_2^* - z_2^*)} - e^{-\frac{3}{4}(3y_2^* - x_2^* - z_2^*)}}\right)\}, \quad (2)$$

and the second is the event that $\mathbf{x}$ is perceived to be better than $\mathbf{z}$,

$$\{\mathbb{E}[u(\mathcal{X})|X, Y, Z] > \mathbb{E}[u(\mathcal{Z})|X, Y, Z]\} = \{\epsilon_2 < \frac{3}{2} - \frac{4}{3}\ln\left(\frac{e^{-3x_1^*} - e^{-3z_1^*}}{e^{-\frac{3}{4}(3z_2^* - x_2^* - y_2^*)} - e^{-\frac{3}{4}(3x_2^* - y_2^* - z_2^*)}}\right)\}. \quad (3)$$

In these two events, both fractions inside the logarithm are positive because $z_1^* > x_1^* > y_1^*$ and $y_2^* > x_2^* > z_2^*$. It is clear that both sets are monotonic in the attributes of $\mathbf{x}$; the better the attributes of $\mathbf{x}$ are, the larger the probability that $\mathbf{x}$ is the most preferred. Through a similar rationale, it is intuitive to see in Equation (3) that the event that $\mathbf{x}$ is preferred to $\mathbf{z}$ is monotonically decreasing in $\mathbf{z}$'s attributes.

More subtle is the influence of the attributes of $\mathbf{z}$ on the preference between $\mathbf{x}$ and $\mathbf{y}$. From Equation (2), it is clear that the first attribute of z_1^* does not affect the preference between $\mathbf{x}$ and $\mathbf{y}$, because the first attribute is perceived noiselessly for all. The second attribute is not. The (main component of the) perceived second attribute of $\mathbf{x}$ is $3x_2^* - y_2^* - z_2^*$.[18] Hence the term $-e^{-\frac{3}{4}(3x_2^* - y_2^* - z_2^*)}$ is (the main component of) the posterior

[18] This can be seen in the fact that the posterior belief is $\mathcal{X}_2 \sim \mathcal{N}(\frac{1}{4}(3X_2 - Y_2 - Z_2), \frac{1}{4})$.

utility of $\mathbf{x}$ from the second attribute. A weak attribute level for z_2^* contrasts with that of $\mathbf{x}$, increasing $\mathbf{x}$'s perceived level and its posterior utility level. Therefore, $\mathbf{x}$ appears to be more appealing in the context of an undesirable $\mathbf{z}$. Similarly, such an undesirable $\mathbf{z}$ also increases the posterior utility of $\mathbf{y}$. However, since $y_2^* > x_2^*$, $\mathbf{y}$ is more saturated than $\mathbf{x}$ in the second attribute. Hence, the increase in the perceived levels benefits $\mathbf{x}$ more. Mathematically, both the posterior utility of $\mathbf{y}$ and of $\mathbf{x}$ from the second attribute increase as z_2^* decreases, but their gap

$$e^{-\frac{3}{4}(3x_2^*-y_2^*-z_2^*)} - e^{-\frac{3}{4}(3y_2^*-x_2^*-z_2^*)} = -e^{-\frac{3}{4}(3y_2^*-x_2^*-z_2^*)} - \left(-e^{-\frac{3}{4}(3x_2^*-y_2^*-z_2^*)}\right)$$
$$= \left(-e^{-\frac{3}{4}(3y_2^*-x_2^*)} - (-e^{-\frac{3}{4}(3x_2^*-y_2^*)})\right)\exp(\frac{3}{4}z_2^*)$$

decreases. Therefore, from Equation (2), a low z_2^* benefits $\mathbf{x}$ more, causing $\mathbf{x}$ to be preferred to $\mathbf{y}$.

To show that the compromise effect occurs, we take the limit as $z_1^* \to x_1^*$ from the right and see from Equation (3) that $\mathbf{x}$ is perceived to be better than $\mathbf{z}$ with probability approaching 1; i.e., $\Pr\left(\left\{\mathbb{E}[u(\mathcal{X})|X,Y,Z] > \mathbb{E}[u(\mathcal{Z})|X,Y,Z]\right\}\right) \to 1$ as $z_1^* \searrow x_1^*$. Moreover, for a small enough z_2^*, the event in Equation (2) becomes a superset of the event in Equation (1); i.e., $\Pr\left(\left\{\mathbb{E}[u(\mathcal{X})|X,Y,Z] > \mathbb{E}[u(\mathcal{Y})|X,Y,Z]\right\}\right) > C(\mathbf{x},\{\mathbf{x},\mathbf{y}\})$ for a small enough z_2^*. Therefore, $C(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{z}\}) > C(\mathbf{x},\{\mathbf{x},\mathbf{y}\})$ for inferior enough $\mathbf{z}$. We have just proved the following result.

Proposition 3. *Assume the parametrization in this section. For any* $\mathbf{x},\mathbf{y}$ *with* $x_1^* > y_1^*$ *and* $x_2^* < y_2^*$*, there exists a* $\delta > 0$ *and a* $D \in \mathbb{R}$ *such that for all* $\mathbf{z}$ *with* $z_1^* - x_1^* \in (0,\delta)$ *and* $z_2^* < D$*, the compromise effect holds.*

The result above identifies an important distinction between our model and a large class of models that satisfy monotonicity (also called regularity). This includes the class of all random utility models (see e.g., Block and Marschak (1960) and Falmagne (1978) and section 5 of Rieskamp et al. (2006)). In the random utility framework, the utility of the options $\mathbf{x},\mathbf{y},\mathbf{z}$ are the random variables $U_{\mathbf{x}}, U_{\mathbf{y}}, U_{\mathbf{z}}$, i.e., measurable functions from a probability space to $\mathbb{R}$. The decision maker chooses $\mathbf{x}$ if and only if the event $\{U_{\mathbf{x}} > U_{\mathbf{y}}$ and $U_{\mathbf{x}} > U_{\mathbf{z}}\}$ is realized. A very general random utility model allows $U_{\mathbf{x}}, U_{\mathbf{y}}$, and $U_{\mathbf{z}}$ to be correlated in arbitrary ways. Nonetheless for a random utility model, it always holds that

$$\{U_{\mathbf{x}} > U_{\mathbf{y}}\} \supseteq \{U_{\mathbf{x}} > U_{\mathbf{y}} \text{ and } U_{\mathbf{x}} > U_{\mathbf{z}}\}, \text{ and hence } C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) \geq C(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{z}\}).$$

According to Proposition 3, our model directly violates this property, and hence, it cannot be reinterpreted as a random utility model.

3.5. Remarks on the Parametric Model

The above examples illustrate how intransitivity, j-s reversals and the compromise effect can be explained by the parametric model when some attributes are subject to imperfect perception. Further derivation reveals that, for the same model, these effects can occur under a range of budget sets that are expected from

Table 1 Some choice sets that correspond to some choice effects observed in the behavioral literature.

Observed patterns	Some consistent **x** and **y**	Corresponding empirical interpretations
Intransitivity: there is a $\mathbf{z}$ with $z_1^* < y_1^* < x_1^*$ and $z_2^* > y_2^* > x_2^*$, so that $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) \geq .5$, $C(\mathbf{y},\{\mathbf{y},\mathbf{z}\}) > .5$ and $C(\mathbf{z},\{\mathbf{z},\mathbf{x}\}) > .5$.	$0 < x_1^*$; $-\frac{3}{2} \approx y_1^*$; $0 \approx x_2^*$; $x_2^* < y_2^* \leq 3$.	Between $\mathbf{x},\mathbf{y}$, a smaller $x_1^* - y_1^*$ than $y_2^* - x_2^*$ is enough for $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) \geq \frac{1}{2}$. And between $\mathbf{y},\mathbf{z}$, a smaller $y_1^* - z_1^*$ than $z_2^* - y_2^*$ is enough for $C(\mathbf{y},\{\mathbf{y},\mathbf{z}\}) > \frac{1}{2}$. However, between $\mathbf{x},\mathbf{z}$, $z_2^* - x_2^*$ is now large enough to be consequential. So $C(\mathbf{z},\{\mathbf{z},\mathbf{x}\}) > \frac{1}{2}$.This matches the empirical pattern that a small difference in attribute one is decisive if the difference in attribute two is small. Attribute two is decisive when its difference is large enough (Tversky 1969, Leland 1994).
J-s reversal: for some $\mathbf{x},\mathbf{y}$ where $x_1^* > y_1^*$, $x_2^* < y_2^*$, it holds that $\$(\mathbf{x}) > \$(\mathbf{y})$ in a separate evaluation, and $\$(\mathbf{x}\|\mathbf{x},\mathbf{y}) < \$(\mathbf{y}\|\mathbf{x},\mathbf{y})$ in a joint evaluation.	$0 < x_1^*$; $y_1^* < -\frac{6}{5}$; $x_2^* \approx 0$; $-3y_1^* < y_2^*$.	Here, attribute one is perceived noiselessly but attribute two is not. Since the better perceived attribute is easier to evaluate, having a smaller $x_1^* - y_1^*$ than $y_2^* - x_2^*$ can lead to $\$(\mathbf{x}) > \$(\mathbf{y})$ in separate evaluations. However, in joint evaluations, the comparison improves the perception of attribute two. So a large enough $y_2^* - x_2^*$ can reverse the evaluations and $\$(\mathbf{y}\|\mathbf{x},\mathbf{y}) > \$(\mathbf{x}\|\mathbf{x},\mathbf{y})$. This matches the empirical pattern termed "evaluability hypothesis" in (Hsee et al. 1999).
Compromise Effect: there is a $\mathbf{z}$ with $y_1^* < x_1^* < z_1^*$ and $y_2^* > x_2^* > z_2^*$ so that $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) < C(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{z}\})$	$y_1^* < x_1^*$; $x_2^* < y_2^*$.	As in the model, when $\mathbf{x}$ is better than $\mathbf{y}$ in attribute one and $\mathbf{y}$ is better in attribute two, introducing an option $\mathbf{z}$ that is best in attribute one and worst in attribute two makes $\mathbf{x}$ a compromise. This can increase the choice probability of $\mathbf{x}$ (Simonson 1989).

the behavioral literature. Table 1 summarizes some of these budget sets and how they can correspond to observations from the behavioral literature.

In addition to explaining the empirical observations, our parametric model can be used to estimate preferences and then to predict choice probabilities for new choice sets. For example, let the utility function be additively exponential $u(x) := u(x_1, x_2) = -e^{\gamma x_1} - e^{\rho x_2}$ where $\gamma, \rho < 0$ are preference parameters. The noise is $\epsilon \sim N\left(0, \begin{bmatrix} 1/t_1^2 & 0 \\ 0 & 1/t_2^2 \end{bmatrix}\right)$ with parameters $t_1, t_2 \in (0, \infty]$, one of which is potentially infinite. Under this parametrization, the parameters can be estimated easily from choice data. For example, in our parametrization, the choice probability for any binary problem is given analytically in Lemma 1.

LEMMA 1. *For any* $\mathbf{x},\mathbf{y}$ *where* $x_1^* > y_1^*$ *and* $y_2^* > x_2^*$, *the parametric model in this subsection gives* $C(\mathbf{x},\{\mathbf{x},\mathbf{y}\}) = \Phi\left(\theta(\gamma,\rho,\mathbf{x}^*,\mathbf{y}^*,t)\right)$, *where* Φ *is the standard normal c.d.f. and* $\theta(\gamma,\rho,\mathbf{x}^*,\mathbf{y}^*,t)$ *is defined as*

$$\theta := \frac{1}{\sqrt{\left(\frac{\rho\sqrt{t_2}}{2+t_2^2}\right)^2 + \left(\frac{\gamma\sqrt{t_1}}{2+t_1^2}\right)^2}} \left[\frac{\gamma^2}{2(2+t_1^2)} - \frac{\rho^2}{2(2+t_2^2)} + \ln\left(\frac{\exp\left(\gamma\frac{(t_1^2+1)y_1^*-x_1^*}{2+t_1^2}\right) - \exp\left(\gamma\frac{(t_1^2+1)x_1^*-y_1^*}{2+t_1^2}\right)}{\exp\left(\rho\frac{(t_2^2+1)x_2^*-y_2^*}{2+t_2^2}\right) - \exp\left(\rho\frac{(t_2^2+1)y_2^*-x_2^*}{2+t_2^2}\right)}\right)\right].$$

When an attribute becomes noiselessly perceptible (i.e. $t_1 \to \infty$), the above Lemma reduces to Equation 1. As seen previously, an $\mathbf{x}$ with better attributes results in a higher θ and hence higher $C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\})$, and vice versa. Moreover, because $x_1^* > y_1^*$, and γ is the preference parameter in the first attribute, a larger γ^2 implies that the first attribute is more decisive, and hence the agent is more likely to choose $\mathbf{x}$.

As the Lemma specifies the choice probabilities in terms of parameters, it can be used to estimate exponential utility functions when there are observations from different menus. After the parameters are estimated, the model can be used to predict choice probabilities for new menus. When doing so, we adopt an implicit assumption similar to Koszegi and Szeidl (2013) that to maintain empirical identifiability and avoid excessive degrees of freedom, the definitions and measurements of the attributes must be determined *before* fitting the model to data. They should not be free parameters but part of the data that the model seeks to explain.[19]

Although the expression in the Lemma can be useful for experimenters, the agent in the model does not evaluate this complicated algebra before making her choice. She simply chooses the choice item that maximizes her expected utility while being unaware of the choice probabilities her actions generate.

4. General Results

The previous section shows that one simple parametric setting can explain and predict several contextual effects. These results are *not* outcomes of parametric flexibility. In contrast, the next subsection shows that the model is robust in the sense that contextual effects will always occur in some correctly designed choice problems. In other words, there is no need to use a different set of parameters to explain each context effect.

Apart from the correctly designed choice problems, contextual effects are rarely observed in many other choice problems. There, the observed choices are usually more "rational". Subsection 4.2 shows that this is in accordance with our model. The agent's choice conforms to classical choice theory for a class of choice problems in which contextual effects are not empirically observed. We then present some intuitive regularity conditions that our model always satisfies.

4.1. The Decoy Choice Pattern

Section 3.4 shows that the model accounts for the compromise effect. Through a similar mechanism, our model also captures two other effects in Figure 3.[20] The *phantom decoy* effect (Pratkanis and Farquhar 1992) occurs in situations in which $\mathbf{z}$ is positioned near the area P. The phantom alternative is better than $\mathbf{x}$ in the

[19] While it is easier to follow this procedure in marketing experiments in which the attributes of each object are specified by the experimenter, it is sometimes difficult to include other relevant attributes in real life decision-making processes. For example, when shopping (online or in person), individuals may base their decisions on attributes that are not listed on the product descriptions. For instance, decisions may be made based on the retailer's customer service, which is usually not listed on product labels. Hence it is difficult to account for these influences.

[20] Here, we omit the formal proofs in the interest of space. The proofs are similar to the intuitive argument in the apartment hunting example from the introduction. By looking at the contrast effect in one of the attributes, the new option $\mathbf{z}$ changes the perceptions of $\mathbf{x}$ and $\mathbf{y}$ in that attribute and affects the relative expected utility.

first attribute and no worse than $\mathbf{x}$ in the second. Additionally, it is worse than $\mathbf{y}$ in the second attribute. In experiments, the subjects are told that such a $\mathbf{z}$ is unavailable; hence, the subject has to choose from $\{\mathbf{x}, \mathbf{y}\}$. Empirically, the phantom decoy increases the frequency of choosing $\mathbf{x}$.[21] The *attraction effect* (Huber et al. 1982) corresponds to the effect of introducing a third option $\mathbf{z}$ in or near the region A in Figure 3. In general, $\mathbf{z}$ needs to be inferior to $\mathbf{x}$ in the second attribute, and no better in the first. In addition, $\mathbf{z}$ needs to be better than $\mathbf{y}$ in the first attribute. Empirically, such a third option itself is rarely chosen but increases the choice frequency of $\mathbf{x}$. Both effects violate monotonicity.

Because our model predicts these three effects through a similar channel, it suggests that there could be some commonality among the effects, as argued by Highhouse (1996). To summarize, start with a binary choice problem in which $\mathbf{x}$ is better than $\mathbf{y}$ in the first attribute but $\mathbf{y}$ is better in the second, as shown in Figure 3. A third object $\mathbf{z}$ in the lower right corner of Figure 3 generally increases the choice probability of $\mathbf{x}$. Due to symmetry, it is also true empirically that if, instead of $\mathbf{z}$, a third object $\mathbf{w}$ lies in the upper left corner of the same figure, the choice probability of $\mathbf{y}$ will increase (i.e., there will be a compromise effect where $\mathbf{y}$ is the compromise option). These empirical effects share the common feature that $\mathbf{z}$ and $\mathbf{w}$ are either unavailable (phantom decoys) or rarely chosen (as in the compromise effect or attraction effect). Therefore, one can reasonably infer that both the attraction effect and the compromise effect remain qualitatively unchanged when the third option is unavailable. To conceptualize these observations, there exist some $\mathbf{w}$ and $\mathbf{z}$ for which the difference $\mathbf{z}^* - \mathbf{w}^*$ points towards the lower-right half plane, such that the *unavailable third option* $\mathbf{w}$ increases the choice probability of $\mathbf{y}$ whereas the *unavailable third option* $\mathbf{z}$ increases the choice probability of $\mathbf{x}$. We call this comparative statics the *decoy choice pattern*.

Definition 1. A choice behavior is said to display the *decoy choice pattern* if there exists a vector $\Delta \in \mathbb{R}^2$ with $\Delta_1 > \Delta_2$, such that for any $\mathbf{x}, \mathbf{y}$ with attributes in $\mathbb{R}^2$ satisfying $x_1^* > y_1^*$, $x_2^* < y_2^*$ the inequality $C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, (\mathbf{z})\}) > C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, (\mathbf{w})\})$ holds whenever $\mathbf{z}^* = \mathbf{w}^* + \lambda\Delta$ for some $\lambda > 0$.

Our model predicts the decoy choice pattern under the general class of preferences and prior-signal distributions as described in Section 2.

Theorem 1. *Any normal-Bayesian agent with a standard preference and imperfect perception displays the decoy choice pattern.*

Observe that Theorem 1 is a sufficiency result. Intuitively, it states that for some $\mathbf{z}^*$ to the right or to the bottom of $\mathbf{w}^*$, such a $\mathbf{z}$ affects the choice probability of $\mathbf{x}$ positively compared to $\mathbf{w}$. Another interesting implication of the theorem is that the attraction effect and the compromise effect should still exist even when $\mathbf{z}$ is a "phantom" option. These predictions are possibly quantitatively too strong in reality but can still be qualitatively reasonable since $\mathbf{z}$ is rarely chosen in experiments. This is a prediction unique to our model, as other models usually do not consider phantom options.

[21] See e.g. Pratkanis and Farquhar (1992), Highhouse (1996), Pettibone and Wedell (2000), Pettibone and Wedell (2007) and Hedgcock et al. (2009).

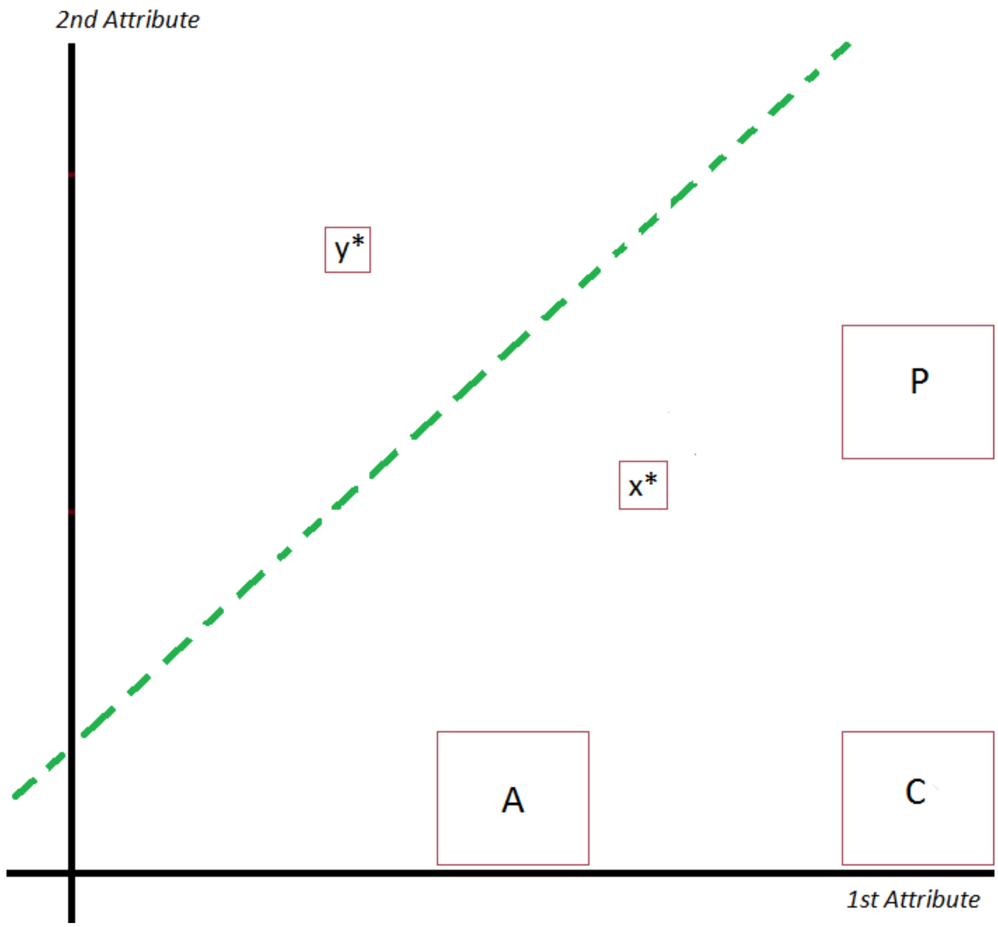


Figure 3 **Areas for the phantom decoy effect (P), the compromise effect (C) and the attraction effect (A)**

4.2. Choice under Dominance

We have seen previously that when there is a trade-off between alternatives — i.e., some alternatives are better in the first attribute while others are better in the second — contextual choices can arise in the model. A natural question is what the model predicts when such a trade-off is absent. Intuitively, if we are given two alternatives $\mathbf{x}$ and $\mathbf{z}$ where $\mathbf{z}^* > \mathbf{x}^*$, a rational agent should always choose $\mathbf{z}$ due to the monotonicity of the utility function.[22] The prediction of our model fits this intuition. Since the error ϵ in perception is the same for both $\mathbf{x}$ and $\mathbf{z}$, the perturbed signal $X = \epsilon + \mathbf{x}^*$ and $Z = \epsilon + \mathbf{z}^*$ preserves the inequality: $Z > X$. The Bayesian agent in our model can hence correctly infer the inequality and choose optimally.

Theorem 2. *For any $\{\mathbf{x}, \mathbf{z}\}$ with $\mathbf{x}^*, \mathbf{z}^* \in \mathbb{R}^2$, a normal-Bayesian agent with standard preference and imperfect perception chooses $\mathbf{z}$ with probability 1 if $\mathbf{z}^* > \mathbf{x}^*$.*

It is clear that the above theorem also predicts the following intuitive choice effect described and observed in Tversky (1972) and Tversky and Russo (1969). Consider an individual who is choosing between a trip

[22] The vector inequality $\mathbf{z}^* > \mathbf{x}^*$ means $z_1^* \geq x_1^*$ and $z_2^* \geq x_2^*$ with at least one inequality being strict.

to Paris ($\mathbf{x}$) and a trip to Rome ($\mathbf{y}$). If she is interested in seeing both places and does not have a strong preference for one over the other, her choice probability for $\mathbf{x}$ would be roughly $1/2$. Now if we offer the individual a new choice problem with two alternatives, a trip to Paris ($\mathbf{x}$) and a trip to Paris plus a \$5 bonus ($\mathbf{z}$), he would not hesitate to choose $\mathbf{z}$ with the extra five dollars. In other words, choosing $\mathbf{z}$ over $\mathbf{x}$ is of probability ≈ 1. However, if we offer her a third choice problem that consists of $\mathbf{y}$ and $\mathbf{z}$, it is intuitive that the choice probability for either one will still be roughly $1/2$.

Another implication of the above theorem is that transitivity holds with overwhelming probability among options that dominate each other. Therefore, one prediction of our model is that violation of weak stochastic transitivity can happen only when the options do not dominate each other. Likewise, our model also predicts that contextual effects do not occur when there is a relationship of dominance in the attributes among all the alternatives.

The proof of the following result is immediate.

COROLLARY 1. *Suppose* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *have attributes* $\mathbf{x}^* > \mathbf{y}^* > \mathbf{z}^*$, *then* $1 = C(\mathbf{x}, \{\mathbf{x}, \mathbf{y}\}) = C(\mathbf{y}, \{\mathbf{y}, \mathbf{z}\}) = C(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) > 1/2$.

When there is only one attribute, the assumption of this Corollary holds. Therefore we predict that contextual effects only occur when there are two or more attributes.

Theorem 2 can also be generalized to the following statement. When $S = \{\mathbf{x}^1, \ldots, \mathbf{x}^n\}$ is the choice set involving multiple options, if $\mathbf{x}^i$ is dominated in the set S, then $C(\mathbf{x}^i, S) = 0$. In other words, objects are chosen with positive probability only when they are on the "attribute possibility frontier". This is a desirable regularity condition that our model satisfies, and it rules out many other types of irregular choice behaviors.[23]

4.3. Relaxing the Assumption of Perfect Correlations

Imperfect perception assumes that the signals across options have the same noise. In other words, the noise is perfectly correlated across the alternatives. The strength of this assumption simplifies the notation and derivations. However, it is not necessary and can be weakened. For example, when the noise is positively correlated across options, the quantitative properties of our model still hold approximately. In this case, from the decision maker's perspective, it means that she does not have to believe that the noise is identical across options. It suffices for her to believe that only a component of the noise is common, so that the noise is not too different across alternatives.

For instance, when there are alternatives $\{\mathbf{x}, \mathbf{y}\}$ with signals X and Y, let the noise be $\epsilon_x = X - \mathbf{x}^*$ and $\epsilon_y = Y - \mathbf{y}^*$. Let us use μ to denote the posterior belief of $(\mathcal{X}, \mathcal{Y})|X, Y$ when the noise is perfectly correlated, i.e. $\epsilon_x = \epsilon_y$. Similarly, μ_a is used to denote the posterior belief when perfect correlation *does*

[23] It is also one of the distinctions between our model and Natenzon (2019). In his model, a dominated object $\mathbf{x}$ with $\mathbf{x}^* < \mathbf{y}^*$ can still be chosen with a probability significantly greater than 0.

not hold. Suppose that under belief μ, the agent chooses $\mathbf{x}$, i.e., $\mathbb{E}_\mu[u(\mathcal{X})] > \mathbb{E}_\mu[u(\mathcal{Y})]$. Note that when μ_a is close enough to μ, it also holds that $\mathbb{E}_{\mu_a}[u(\mathcal{X})] > \mathbb{E}_{\mu_a}[u(\mathcal{Y})]$ and so $\mathbf{x}$ is also chosen under belief μ_a. Therefore, due to this continuity, one can locally relax the assumption and allow $\epsilon_x \neq \epsilon_y$, and at the same time the resulting μ_a would be close enough to μ for the model predictions to be quantitatively similar.

Formally, the following convergence result shows that when the correlation across alternatives is high enough, the posterior μ_a is close enough to μ. For commonly used utility functions, such closeness is sufficient to maintain the choice decisions in a given problem.

Proposition 4. *Let the prior for each option be $\mathcal{N}(0,\Omega)$. For any n options with realized signals $X^1,\dots,X^n$, let the noise for each signal be ϵ^i. Suppose $(\epsilon^1,\dots,\epsilon^n)' \sim \mathcal{N}(0,\Sigma_a)$ for some positive definite Σ_a, and the resulting posterior belief be μ_a. Denote by Σ the $2n \times 2n$ matrix*

$$\Sigma = \begin{bmatrix} T^{-1} & T^{-1} & \dots & T^{-1} \\ \vdots & \vdots & \dots & \vdots \\ T^{-1} & T^{-1} & \dots & T^{-1} \end{bmatrix},$$

and by μ the posterior belief when $\epsilon^i = \epsilon^j$ almost surely. Then $\mu_a(\mathcal{X}^i)$ for each i is normally distributed and weakly converges to $\mu(\mathcal{X}^i)$ as $\Sigma_a \to \Sigma$.

As seen in Section 2, positively correlated noise is empirically plausible. Therefore, we impose this assumption as it is theoretically desirable for a relatively restrictive assumption to have strong explanatory power. However, it is also of natural curiosity to think about other assumptions, such as those of negatively correlated signals. Because the mathematical nature of negative correlations forbids a direct such generalization, we do not extensively explore more exotic generalizations in this paper due to limited scope. The degree of the negative correlation is limited by the number of signals (and hence the number of options) at hand. When there are two signals X and Y, they can be perfectly negatively correlated, or $corr(X,Y) = -1$. When there are three signals, X, Y, and Z, if $corr(X,Y) = -1$ and $corr(Y,Z) = -1$, then it implies $corr(X,Z) = 1$. If they all have to be negatively correlated, the possible limit is $corr(X,Y) = corr(X,Z) = corr(Y,Z) = -1/2$. As the number of signals increases, the negative correlation between the signals has to diminish to zero. If we want to maintain, as in our model, that the joint distribution of (X,Y) does not change from choice set to choice set, then a model in which all options have negatively correlated signals is not feasible.

4.4. A Limiting Noise Structure

As shown in Proposition 3, our model does not satisfy monotonicity, a fundamental property of all random utility models. Despite this difference, one interesting question is whether such non-monotonic predictions disappear for some limiting parameters of our model. For example, if the noise in the signal goes to zero, does our model converge to some well-known model? The following shows that as the noise term becomes small, our model approximates the well-known conditional probit model of Hausman and Wise (1978).

Because this conditional probit model can explain the similarity effect, a corollary of this subsection is that our model can also explain the similarity effect.

Again, we restrict our discussion to exponential utility functions so that $u(x_1,x_2) = -e^{\gamma x_1} - e^{\rho x_2}$. Given a finite choice set $S = \{\mathbf{x}^1, \dots \mathbf{x}^n\}$, the posterior belief over the i-th alternative under imperfect perception is

$$\mathcal{X}^i | X^1, \dots X^n \sim \mathcal{N}\left((T + n\Omega^{-1})^{-1} \left(TX^i + n\Omega^{-1}X^i - \sum_{j=1}^{n} \Omega^{-1}X^j \right), (T + n\Omega^{-1})^{-1} \right).$$

When the variance of the noise converges to zero, i.e. $T^{-1} \to \mathbf{0}$, the posterior belief $\mathcal{X}^i | X^1, \dots X^n$ is approximately $\mathcal{N}(X^i, T^{-1}) = \mathcal{N}(\mathbf{x}^{i*} + \epsilon, T^{-1})$. When the utility function is smooth enough near $\mathbf{x}^{i*}$, we approximate the expected utility by using the utility of the expected attributes

$$\mathbb{E}[u(\mathcal{X}) | X^1, \dots X^n] \approx u(\mathbf{x}^{i*} + \epsilon)$$

which is already a random utility model. Under exponential utility, this approximates the Hausman and Wise (1978) model,

$$u(\mathbf{x}^{i*} + \epsilon) = -e^{\gamma(x_1^{i*} + \epsilon_1)} - e^{\rho(x_2^{i*} + \epsilon_2)} \approx u_1(x_1^{i*}) + u_2(x_2^{i*}) + \beta_1 u_1(x_1^{i*}) + \beta_2 u_2(x_2^{i*}),$$

where we have used the first-order approximation of $\mathbf{x}^{i*}$ with the notation that $u_1(x_1) = -e^{\gamma x_1}, u_2(x_2) = -e^{\rho x_2}$ and $\beta_1 = \gamma \epsilon_1, \beta_2 = \rho \epsilon_2$. It is clear that the form of the approximation coincides with equation (3.6) in Hausman and Wise (1978).

4.5. The Link with Reference-dependent Theories

Many contextual effects can also be explained with reference-dependent utility models. It is therefore of interest to explore how our model relates to this class of models. As we take the limit of our prior variance to zero, the limiting case becomes choice-equivalent to a reference-dependent model.

In the limit as $\Omega \to 0$, direct calculation shows that the limiting posterior becomes a Dirac measure, i.e. $\mathcal{X} | X^1, \dots X^n = X^i - \frac{1}{n} \sum_{\mathbf{x}^j \in S} X^j$ with probability 1. Hence the utility of $\mathbf{x}^i$ becomes

$$u\left(X^i - \frac{1}{n} \sum_{\mathbf{x}^j \in S} X^j \right) = u\left(\mathbf{x}^{i*} + \epsilon - \frac{1}{n} \sum_{\mathbf{x}^j \in S} (\mathbf{x}^{j*} + \epsilon) \right) = u(\mathbf{x}^{i*} - \bar{\mathbf{x}}),$$

where $\bar{\mathbf{x}} = \frac{1}{n} \sum_{\mathbf{x}^j \in S} \mathbf{x}^{j*}$. Clearly, this is a reference-dependent utility function in which the reference point for a choice set S is the average attribute in S. Nonetheless, this limiting reference-dependent utility cannot be directly applied to explain evaluation outcomes such as j-s reversal. For any two alternatives $\mathbf{x}$ and $\mathbf{y}$, the separate evaluations are equal:

$$u(\mathbf{x}^* - \frac{1}{1} \sum_{\mathbf{x} \in \{\mathbf{x}\}} \mathbf{x}^*) = u(\mathbf{0}) = u(\mathbf{y}^* - \frac{1}{1} \sum_{\mathbf{y} \in \{\mathbf{y}\}} \mathbf{y}^*).$$

Therefore, strict evaluation reversals never occur. Similarly, in the reference-dependent model of Tserenjigmid (2019), the reference point is the minimum vector of all the attributes, which also normalizes an object's utility to $u(\mathbf{0})$ in separate evaluations. Therefore, j-s reversal cannot occur. In general, the decision maker is equally happy when there is only one alternative available regardless of its attribute levels when the reference point normalizes it to the origin.

4.6. Comparing with Existing Reference-dependent Theories

4.6.1. Comparing with Koszegi and Rabin (2006)

The model of Koszegi and Rabin (2006) satisfies the property Expansion (defined below) which can be violated in our model.

Definition 2 (Expansion). If $\mathbf{y}$ is chosen in a choice set D, and $\mathbf{y}$ is also chosen in a choice set D', then $\mathbf{y}$ is chosen in $D \cup D'$.

A proof of Koszegi and Rabin (2006) satisfying this property can be found in Proposition 2 and Observation 3 in Freeman (2017). However, Expansion is not qualitatively satisfied by our model and the compromise effect.

To show this in our setting, we first make the following interpretation: define "$\mathbf{y}$ is chosen in a set D" to be "$\mathbf{y}$ is chosen with highest probability in D".[24] Suppose for some $x_1^* > y_1^*$ and $x_2^* < y_2^*$, both $\mathbf{x}$ and $\mathbf{y}$ are chosen in $\{\mathbf{x}, \mathbf{y}\}$. Suppose there is a $\mathbf{z}$ that satisfies $z_1^* > x_1^* > y_1^*$ and $z_2^* < x_2^* < y_2^*$ with z_2^* inferior enough. Our Lemma 1 shows $\mathbf{y}$ would be chosen over $\mathbf{z}$ in $\{\mathbf{z}, \mathbf{y}\}$. Now Expansion requires that $\mathbf{y}$ is chosen in $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\} = \{\mathbf{x}, \mathbf{y}\} \cup \{\mathbf{z}, \mathbf{y}\}$. However, our model allows the compromise effect in which $\mathbf{x}$ is chosen over $\mathbf{y}$ in $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Therefore, the property Expansion is not satisfied.

4.6.2. Comparing with Bordalo et al. (2013)

The model of Bordalo et al. (2013) is very general and does not satisfy our Theorem 1. For instance, for some $\mathbf{x}$ and $\mathbf{y}$ where $x_1^* > y_1^*$ and $x_2^* < y_2^*$, let $\mathbf{w}$ be an option dominated by both $\mathbf{x}$ and $\mathbf{y}$ in both attributes. There exists such a $\mathbf{w}$ so that changing $\mathbf{w}$ to any other $\mathbf{z}$ can only increase the utility of $\mathbf{y}$ relative to $\mathbf{x}$. This violates our Theorem 1 which states that there exists some $\mathbf{z}$ that can induce a higher choice probability of $\mathbf{x}$ over $\mathbf{y}$. We provide an explicit example of such violation in Appendix A.5.

4.6.3. Comparing with Ok et al. (2015)

Ok et al. (2015) characterized their reference dependent model in their Theorem 1. And their model satisfies the property No-Cycle (defined below), which can be violated in our model.

Definition 3 (No-Cycle). For every $\mathbf{x}, \mathbf{y}, \mathbf{z}$, if $\mathbf{x}$ is chosen in $\{\mathbf{x}, \mathbf{y}\}$, and $\mathbf{y}$ is chosen in $\{\mathbf{y}, \mathbf{z}\}$, then $\mathbf{x}$ is chosen in $\{\mathbf{x}, \mathbf{z}\}$.

This property is qualitatively violated by our model, as our model allows intransitivity. This can be readily seen with the interpretation that "$\mathbf{y}$ is chosen in a set D" is defined as "$\mathbf{y}$ is chosen with highest probability in D".

[24] It can also be shown under other qualitatively similar definition with small changes to the argument.

4.6.4. Comparing with Tserenjigmid (2019)

As mentioned in Section 4.5, the model of Tserenjigmid (2019) does not explain the j-s reversal. Tserenjigmid (2019) defines the utility of an option $\mathbf{x}$ in the choice set D as $u(\mathbf{x}|D) := f(\mathbf{x}^* - \underline{x}_D)$ where the reference point $\underline{x}_D := (\min_{\mathbf{y}\in D} y_1^*, \min_{\mathbf{y}\in D} y_2^*)$, and f is some fixed increasing concave function. Therefore, for any $\mathbf{x}$, it holds that $u(\mathbf{x}|\{\mathbf{x}\}) = f(0)$, and hence the model cannot be applied to explain the j-s reversal.

5. Discussion and Conclusion

This paper presents a choice model with fixed underlying preferences. Through noisy attribute perception, the model generically predicts several contextual effects for the appropriate choice problems in the attribute space. The information structure in our model is exogenously assumed to be dependent on the context. While there are many ways to model general information structures, we adopt one that is minimal with a rather specific mathematical form for its psychological plausibility and mathematical simplicity. Such an informational friction generates both benefits and limitations.

The benefits are twofold. First, the informational friction is restrictive: there is a high positive correlation across alternatives. Although a restrictive assumption generally reduces explanatory power, it can be theoretically desirable when it rules out certain unobserved choice behaviors. For example, we predict the absence of intransitivity when the alternatives dominates each other in attributes, a feasible outcome that is not empirically observed. Second, our approach shows that with minimally restrictive informational frictions, the canonical Bayesian utility maximization can also provide good explanatory power for several contextual effects. This implies that within Bayesian decision theory, each type of choice effect may potentially correspond to a type of informational friction. This correspondence may be of interest in future research.

Nonetheless, there are also several limitations. This simple friction can be too restrictive in some respects while being too general in others. Although the model qualitatively matches the phantom decoy effect, it can be too restrictive since it quantitatively makes the same prediction for unavailable alternatives as for those that are in the choice set. For example, it makes the same prediction in choice probabilities for the attraction effect when the dominated option is present but unavailable. This cannot quantitatively explain the observations in Sivakumar and Cherian (1995) that the choice probability of the target is significantly reduced following the removal of the dominated option.[25] This limitation comes from the exogenous information structure and a potential future project is to endogenize the information structure and capture such empirical findings.

At the same time, our model is potentially too general because it does not rule out some unintuitive predictions. For instance, when the correlations r and R are not restricted, the model permits j-s reversal for a dominated option. Although $\mathbf{x}^* < \mathbf{z}^*$ implies that $\$(\mathbf{x}|\mathbf{x},\mathbf{z}) < \$(\mathbf{z}|\mathbf{x},\mathbf{z})$, it does not imply that $\$(\mathbf{x}) < \$(\mathbf{z})$. This counterintuitive prediction is allowed when the two attributes are negatively correlated in the prior.

[25] Although the probability does not fully recover to the level when the dominated option was never shown.

When $z_1^* > x_1^*$ and $z_2^* = x_2^*$, in separate evaluations, $\mathcal{Z}_2|Z$ can be smaller than $\mathcal{X}_2|X$ in distribution. When the agent values attribute two more intensely, the model predicts that $\$(\mathbf{x}) > \$(\mathbf{z})$.

In addition to our explanation, other mechanisms are also likely at play in reality. Imagine a choice problem with many options; an agent is asked to rank the options or is asked to choose one from each pair. Our model implies that these two tasks yield consistent outcomes. This is because in our model the agent always learns all available information in a choice problem, and there is no capacity constraint to her learning. This prevents our model from explaining some empirical phenomena (such as the choice overload (Iyengar and Lepper 2000)) in which the learning costs are the main driver. For these choice effects, a more suitable model would likely cover endogenous attention and information acquisition. See Guo (2016) for one such model for explaining choice overload.

Appendix A: Proofs

A.1. Proof of Lemma

Proof of Lemma 1 We directly calculate the expected utility

$$\begin{aligned}\mathbb{E}[u(\mathcal{X})|X,Y] =& \mathbb{E}\left[-e^{\gamma\mathcal{X}_1} - e^{\rho\mathcal{X}_2}|X,Y\right]\\ =& -\exp\left(\gamma\frac{(t_1^2+1)X_1 - Y_1}{2+t_1^2} + \gamma^2\frac{1}{2(2+t_1^2)}\right) - \exp\left(\rho\frac{(t_2^2+1)X_2-Y_2}{2+t_2^2} + \rho^2\frac{1}{2(2+t_2^2)}\right)\\ =& -\exp\left(\gamma\frac{(t_1^2+1)x_1^* - y_1^* + t_1^2\epsilon_1}{2+t_1^2} + \gamma^2\frac{1}{2(2+t_1^2)}\right) - \exp\left(\rho\frac{(t_2^2+1)x_2^* - y_2^* + t_2^2\epsilon_2}{2+t_2^2} + \rho^2\frac{1}{2(2+t_2^2)}\right)\end{aligned}$$

where the second equality is due to the normally distributed exponents. The third equality is due to the identities $\mathbf{x}^* + \epsilon = X$, $\mathbf{y}^* + \epsilon = Y$. Similarly,

$$\mathbb{E}[u(\mathcal{Y})|X,Y] = -\exp\left(\gamma\frac{(t_1^2+1)y_1^* - x_1^* + t_1^2\epsilon_1}{2+t_1^2} + \gamma^2\frac{1}{2(2+t_1^2)}\right) - \exp\left(\rho\frac{(t_2^2+1)y_2^* - x_2^* + t_2^2\epsilon_2}{2+t_2^2} + \rho^2\frac{1}{2(2+t_2^2)}\right)$$

Hence given $\mathbf{x}^*, \mathbf{y}^*$ and ϵ, the agent would choose $\mathbf{x}$ over $\mathbf{y}$ iff $\mathbb{E}[u(\mathcal{X})|X,Y] > \mathbb{E}[u(\mathcal{Y})|X,Y]$. Suppose $x_1^* > y_1^*$ and $y_2^* > x_2^*$, then we see that $\mathbf{x}$ is chosen over $\mathbf{y}$ iff

$$\exp\left(\frac{\gamma^2}{2(2+t_1^2)} - \frac{\rho^2}{2(2+t_2^2)}\right)\frac{\exp\left(\gamma\frac{(t_1^2+1)y_1^* - x_1^*}{2+t_1^2}\right) - \exp\left(\gamma\frac{(t_1^2+1)x_1^* - y_1^*}{2+t_1^2}\right)}{\exp\left(\rho\frac{(t_2^2+1)x_2^* - y_2^*}{2+t_2^2}\right) - \exp\left(\rho\frac{(t_2^2+1)y_2^* - x_2^*}{2+t_2^2}\right)} \geq \exp\left(\frac{\rho t_2^2\epsilon_2}{2+t_2^2} - \frac{\gamma t_1^2\epsilon_1}{2+t_1^2}\right). \tag{†}$$

Since $x_1^* > y_1^*$ and $y_2^* > x_2^*$, we can take the natural log on both sides of (†) to obtain the following equivalent condition

$$\frac{\gamma^2}{2(2+t_1^2)} - \frac{\rho^2}{2(2+t_2^2)} + \ln\left(\frac{\exp\left(\gamma\frac{(t_1^2+1)y_1^* - x_1^*}{2+t_1^2}\right) - \exp\left(\gamma\frac{(t_1^2+1)x_1^* - y_1^*}{2+t_1^2}\right)}{\exp\left(\rho\frac{(t_2^2+1)x_2^* - y_2^*}{2+t_2^2}\right) - \exp\left(\rho\frac{(t_2^2+1)y_2^* - x_2^*}{2+t_2^2}\right)}\right) \geq \frac{\rho t_2^2\epsilon_2}{2+t_2^2} - \frac{\gamma t_1^2\epsilon_1}{2+t_1^2}.$$

Note that the RHS follows a normal distribution $\mathcal{N}\left(0, \left(\frac{\rho}{2+t_2^2}\right)^2 t_2 + \left(\frac{\gamma}{2+t_1^2}\right)^2 t_1\right)$. We can standardize both sides by multiplying $1/\sqrt{\left(\frac{\rho\sqrt{t_2}}{2+t_2^2}\right)^2 + \left(\frac{\gamma\sqrt{t_1}}{2+t_1^2}\right)^2}$. Hence $\mathbf{x}^*$ is chosen over $\mathbf{y}^*$ iff some standard normal random variable Z is below the threshold θ defined below:

$$\theta(\gamma,\rho,\mathbf{x}^*,\mathbf{y}^*,t) := \frac{1}{\sqrt{\left(\frac{\rho\sqrt{t_2}}{2+t_2^2}\right)^2 + \left(\frac{\gamma\sqrt{t_1}}{2+t_1^2}\right)^2}}\left[\frac{\gamma^2}{2(2+t_1^2)} - \frac{\rho^2}{2(2+t_2^2)} + \ln\left(\frac{\exp\left(\gamma\frac{(t_1^2+1)y_1^* - x_1^*}{2+t_1^2}\right) - \exp\left(\gamma\frac{(t_1^2+1)x_1^* - y_1^*}{2+t_1^2}\right)}{\exp\left(\rho\frac{(t_2^2+1)x_2^* - y_2^*}{2+t_2^2}\right) - \exp\left(\rho\frac{(t_2^2+1)y_2^* - x_2^*}{2+t_2^2}\right)}\right)\right].$$

A.2. Proof of Theorem 1

Proof of Theorem 1 It suffices to show that under our assumptions, for every realization of ϵ the following inequality holds

$$\mathbb{E}[u(\mathcal{X})|X,Y,Z]-\mathbb{E}[u(\mathcal{Y})|X,Y,Z]>\mathbb{E}[u(\mathcal{X})|X,Y,W]-\mathbb{E}[u(\mathcal{Y})|X,Y,W].$$

Conditional on X,Y,W, the posterior for $\mathcal{X}$ is

$$\begin{aligned}
&\Pr(\mathcal{X}|X,Y,W)\\
&\propto \exp\left(-\frac{\mathcal{X}'\Omega^{-1}\mathcal{X}}{2}\right)\exp\left(-\frac{\mathcal{Y}'\Omega^{-1}\mathcal{Y}}{2}\right)\exp\left(-\frac{\mathcal{W}'\Omega^{-1}\mathcal{W}}{2}\right)\exp\left(-\frac{(X-\mathcal{X})'T(X-\mathcal{X})}{2}\right)\times 1_{\{X-\mathcal{X}=Y-\mathcal{Y}=W-\mathcal{W}\}}\\
&\propto \exp\left(-\frac{1}{2}\left[\mathcal{X}'\left(3\Omega^{-1}+T\right)\mathcal{X}-2\left(TX-\Omega^{-1}(Y+W-2X)\right)'\mathcal{X}\right]\right)\\
&\propto \exp\left(-\frac{1}{2}\left(\mathcal{X}-\left(3\Omega^{-1}+T\right)^{-1}\left(TX-\Omega^{-1}(Y+W-2X)\right)\right)'\left(3\Omega^{-1}+T\right)\left(\mathcal{X}-\left(3\Omega^{-1}+T\right)^{-1}\left(TX-\Omega^{-1}(Y+W-2X)\right)\right)\right)
\end{aligned}$$

So we denote the above posterior distribution of $\mathcal{X}|X,Y,W$ by $\mathcal{N}\left(\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon),\hat{\Omega}\right)$, where

$$\begin{aligned}
\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon) &:= \left(3\Omega^{-1}+T\right)^{-1}\left(T\mathbf{x}^*+T\epsilon-\Omega^{-1}(\mathbf{y}^*+\mathbf{w}^*-2\mathbf{x}^*)\right)\\
&= \left(3\Omega^{-1}+T\right)^{-1}\left(TX-\Omega^{-1}(Y+W-2X)\right),\\
\text{and}\quad \hat{\Omega} &:= \left(3\Omega^{-1}+T\right)^{-1}.
\end{aligned}$$

Denote the density of $\mathcal{X}|X,Y,W\sim\mathcal{N}(\mu,\hat{\Omega})$ by $\phi(\mathcal{X}-\mu,\hat{\Omega})$. The posterior expected utility is therefore

$$\begin{aligned}
\mathbb{E}[u(\mathcal{X})|X,Y,W] &= \int_{\mathbb{R}^2} u(\mathcal{X})\times\phi\left(\mathcal{X}-\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon),\hat{\Omega}\right)\mathrm{d}\mathcal{X}\\
&= \int_{\mathbb{R}^2} u(\mathbf{s}+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon))\times\phi\left(\mathbf{s},\hat{\Omega}\right)\mathrm{d}\mathbf{s}.
\end{aligned}$$

Similarly,

$$\mathcal{Y}|X,Y,W\sim\mathcal{N}\left(\mu(\mathbf{y}^*;\mathbf{x}^*,\mathbf{w}^*,\epsilon),\hat{\Omega}\right).$$

Because

$$\begin{aligned}
\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon) :=&\hat{\Omega}\left(T\mathbf{x}^*+T\epsilon-\Omega^{-1}(\mathbf{y}^*+\mathbf{w}^*-2\mathbf{x}^*)\right)\\
=&\mu(\mathbf{y}^*;\mathbf{x}^*,\mathbf{w}^*,\epsilon)-(\mathbf{y}^*-\mathbf{x}^*),
\end{aligned}$$

we have

$$\mathbb{E}[u(\mathcal{Y})|X,Y,W]=\int_{\mathbb{R}^2} u(\mathbf{s}+(\mathbf{y}^*-\mathbf{x}^*)+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon))\times\phi\left(\mathbf{s},\hat{\Omega}\right)\mathrm{d}\mathbf{s}.$$

Recall that $\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon)=\hat{\Omega}T\mathbf{x}^*+\hat{\Omega}T\epsilon-\hat{\Omega}\Omega^{-1}\mathbf{y}^*-\hat{\Omega}\Omega^{-1}\mathbf{w}^*+2\hat{\Omega}\Omega^{-1}\mathbf{x}^*$. Substituting $\mathbf{z}^*:=\mathbf{w}^*+\Delta$ for $\mathbf{w}^*$ we have

$$\begin{aligned}
&\mathbb{E}[u(\mathcal{X})|X,Y,Z]-\mathbb{E}[u(\mathcal{Y})|X,Y,Z]\\
=&\int_{\mathbb{R}^2} u(\mathbf{s}+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{z}^*,\epsilon))\times\phi\left(\mathbf{s},\hat{\Omega}\right)\mathrm{d}\mathbf{s}-\int_{\mathbb{R}^2} u(\mathbf{s}+(\mathbf{y}^*-\mathbf{x}^*)+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{z}^*,\epsilon))\times\phi\left(\mathbf{s},\hat{\Omega}\right)\mathrm{d}\mathbf{s}\\
=&\int_{\mathbb{R}^2}\left[u\left(\mathbf{s}+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon)-\hat{\Omega}\Omega^{-1}\Delta\right)-u\left(\mathbf{s}+(\mathbf{y}^*-\mathbf{x}^*)+\mu(\mathbf{x}^*;\mathbf{y}^*,\mathbf{w}^*,\epsilon)-\hat{\Omega}\Omega^{-1}\Delta\right)\right]\times\phi\left(\mathbf{s},\hat{\Omega}\right)\mathrm{d}\mathbf{s}.
\end{aligned}$$

Since u is standard, $y_1^* < x_1^*$, and $y_2^* > x_2^*$, if $-\hat{\Omega}\Omega^{-1}\Delta \in (-\infty, 0) \times (0, \infty)$, i.e. the second quadrant, then

$$\begin{aligned}
&u\left(\mathbf{s} + \mu(\mathbf{x}^*; \mathbf{y}^*, \mathbf{w}^*, \epsilon) - \hat{\Omega}\Omega^{-1}\Delta\right) - u\left(\mathbf{s} + (\mathbf{y}^* - \mathbf{x}^*) + \mu(\mathbf{x}^*; \mathbf{y}^*, \mathbf{w}^*, \epsilon) - \hat{\Omega}\Omega^{-1}\Delta\right) \\
>&u\left(\mathbf{s} + \mu(\mathbf{x}^*; \mathbf{y}^*, \mathbf{w}^*, \epsilon)\right) - u\left(\mathbf{s} + (\mathbf{y}^* - \mathbf{x}^*) + \mu(\mathbf{x}^*; \mathbf{y}^*, \mathbf{w}^*, \epsilon)\right)
\end{aligned}$$

for all $\mathbf{s}$ and ϵ. When we integrate out $\mathbf{s}$, we have $\mathbb{E}[u(\mathcal{X})|X,Y,Z] - \mathbb{E}[u(\mathcal{Y})|X,Y,Z] > \mathbb{E}[u(\mathcal{X})|X,Y,W] - \mathbb{E}[u(\mathcal{X})|X,Y,W]$ for every realization of ϵ.

Therefore, one sufficient condition is that $-\hat{\Omega}\Omega^{-1}\Delta \in (-\infty, 0) \times (0, \infty)$. If this condition holds, we have $-\hat{\Omega}\Omega^{-1}\Delta = \mathbf{w}$ for some $w_1 < 0$, and $w_2 > 0$. To show the decoy choice pattern, we just need to show that there exists some Δ with $\Delta_1 > \Delta_2$ such that this condition holds.

Recall that we had normalized Ω so that for some $r \in (-1, 1)$,

$$\Omega = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

and the noise has variance

$$T^{-1} = \begin{bmatrix} 1/t_1^2 & R/(t_1 t_2) \\ R/(t_1 t_2) & 1/t_2^2 \end{bmatrix}.$$

We can calculate

$$\Omega^{-1} = \begin{bmatrix} 1/(1-r^2) & -r/(1-r^2) \\ -r/(1-r^2) & 1/(1-r^2) \end{bmatrix} \text{ and } T = \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix} \begin{bmatrix} 1/(1-R^2) & -R/(1-R^2) \\ -R/(1-R^2) & 1/(1-R^2) \end{bmatrix} \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix};$$

It follows that

$$\begin{aligned}
\Delta &= -\Omega\hat{\Omega}^{-1}\mathbf{w} = -\Omega(3\Omega^{-1} + T)\mathbf{w} = -(3I + \Omega T)\mathbf{w} \\
&= -\left(\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix} \begin{bmatrix} 1/(1-R^2) & -R/(1-R^2) \\ -R/(1-R^2) & 1/(1-R^2) \end{bmatrix} \begin{bmatrix} t_1 & 0 \\ 0 & t_2 \end{bmatrix}\right) \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\
&= -\begin{bmatrix} 3 + \frac{t_1^2 - t_1 t_2 r R}{1-R^2} & \frac{t_2^2 r - t_1 t_2 R}{1-R^2} \\ \frac{t_1^2 r - t_1 t_2 R}{1-R^2} & 3 + \frac{t_2^2 - t_1 t_2 r R}{1-R^2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}
\end{aligned}$$

Since $w_1 < 0$, and $w_2 > 0$, the sufficient condition holds when Δ is some positive linear combination of the two vectors

$$\left\{\begin{bmatrix} 3(1-R^2) + t_1^2 - t_1 t_2 r R \\ t_1^2 r - t_1 t_2 R \end{bmatrix}, -\begin{bmatrix} t_2^2 r - t_1 t_2 R \\ 3(1-R^2) + t_2^2 - t_1 t_2 r R \end{bmatrix}\right\}.$$

The decoy choice pattern holds when there exists such a Δ with $\Delta_1 > \Delta_2$. In other words, the decoy choice pattern holds if

$$\begin{cases} 3(1-R^2) + t_1^2 - t_1 t_2 r R > t_1^2 r - t_1 t_2 R \\ \text{or} \\ -(t_2^2 r - t_1 t_2 R) > -(3(1-R^2) + t_2^2 - t_1 t_2 r R), \end{cases} \Leftrightarrow \begin{cases} 3(1-R^2) > (r-1)(t_1^2 + t_1 t_2 R) \\ \text{or} \\ 3(1-R^2) > (r-1)(t_2^2 + t_1 t_2 R). \end{cases}$$

Because $r, R \in (-1, 1)$ and $t_1, t_2 > 0$, it is impossible for both $t_1 + t_2 R < 0$ and $t_2 + t_1 R < 0$ to hold simultaneously. Therefore the decoy choice pattern holds.

A.3. Proof of Theorem 2

Proof of Theorem 2 As before, we start with the Bayesian posterior

$$\begin{aligned}\Pr(\mathcal{X}|X,Z) &\propto \exp\left(-\frac{\mathcal{X}'\Omega^{-1}\mathcal{X}}{2}\right)\exp\left(-\frac{\mathcal{Z}'\Omega^{-1}\mathcal{Z}}{2}\right)\exp\left(-\frac{(X-\mathcal{X})'T(X-\mathcal{X})}{2}\right)\times 1_{\{X-\mathcal{X}=Z-\mathcal{Z}\}}\\ &= \exp\left(-\frac{1}{2}\left[\mathcal{X}'\left(2\Omega^{-1}+T\right)\mathcal{X}-2\left(TX-\Omega^{-1}(Z-X)\right)'\mathcal{X}\dots\right]\right)\\ &\propto \exp\left(-\frac{1}{2}\left(\mathcal{X}-\left(2\Omega^{-1}+T\right)^{-1}\left(TX-\Omega^{-1}(Z-X)\right)\right)'\left(2\Omega^{-1}+T\right)\left(\mathcal{X}-\dots\right)\right)\end{aligned}$$

Therefore, the posterior inference for $\mathbf{x}^*$ is

$$\begin{aligned}\mathcal{X}|X,Z \sim &\mathcal{N}\left(\left(2\Omega^{-1}+T\right)^{-1}\left(TX-\Omega^{-1}(Z-X)\right),\left(2\Omega^{-1}+T\right)^{-1}\right)\\ =&\mathcal{N}\left(\left(2\Omega^{-1}+T\right)^{-1}\left(T\mathbf{x}^*+T\epsilon-\Omega^{-1}(\mathbf{z}^*-\mathbf{x}^*)\right),\left(2\Omega^{-1}+T\right)^{-1}\right)\\ :=&\mathcal{N}\left(\mu(\mathbf{x}^*;\mathbf{z}^*,\epsilon),\hat{\Omega}\right)\end{aligned}$$

Similarly, $\mathcal{Z}|X,Z\sim\mathcal{N}\left(\mu(\mathbf{z}^*;\mathbf{x}^*,\epsilon),\hat{\Omega}\right)$. Observe that they have the same variance, and that

$$\begin{aligned}&\mu(\mathbf{z}^*;\mathbf{x}^*,\epsilon)-\mu(\mathbf{x}^*;\mathbf{z}^*,\epsilon)\\ =&\left(2\Omega^{-1}+T\right)^{-1}\left(T\mathbf{z}^*+T\epsilon-\Omega^{-1}(\mathbf{x}^*-\mathbf{z}^*)\right)-\left(2\Omega^{-1}+T\right)^{-1}\left(T\mathbf{x}^*+T\epsilon-\Omega^{-1}(\mathbf{z}^*-\mathbf{x}^*)\right)\\ =&\mathbf{z}^*-\mathbf{x}^*>0.\end{aligned}$$

Therefore the posterior inference distribution for $\mathbf{z}^*$ is that for $\mathbf{x}^*$ translated by the vector $\mathbf{z}^*-\mathbf{x}^*>0$. Since the standard preference is increasing in both attributes, we have for every $\epsilon\in\mathbb{R}^2$

$$\mathbb{E}[u(\mathcal{X})|X,Z]<\mathbb{E}[u(\mathcal{Z})|X,Z].$$

Hence the rational agent chooses $\mathbf{z}$ over $\mathbf{x}$ with probability 1.

A.4. Proof of Proposition 4

Proof of Proposition 4

Define $\tilde{X}=(X^1,\dots,X^n)$, and $\tilde{\mathcal{X}}=(\mathcal{X}^1,\dots,\mathcal{X}^n)$. Let us use $\phi_k(.,A)$ to denote the density of the k-dimensional normal distribution $\mathcal{N}(0,A)$. Under Σ_a, the posterior distribution for X^1 is

$$\Pr(\mathcal{X}^1|X^1,\dots,X^n)=\frac{\int\prod_{j=1}^n\phi_2(\mathcal{X}^j,\Omega)\times\phi_{2n}\left((\tilde{X}-\tilde{\mathcal{X}}),\Sigma_a\right)\mathrm{d}\mathcal{X}^2\times\dots\times\mathrm{d}\mathcal{X}^n}{\int\prod_{j=1}^n\phi_2(\mathcal{X}^j,\Omega)\times\phi_{2n}\left((\tilde{X}-\tilde{\mathcal{X}}),\Sigma_a\right)\mathrm{d}\tilde{\mathcal{X}}}$$

It is easy to see this is a normal distribution, which establishes the claim of the normality of the posterior. Let v be any bounded continuous function. Then

$$\mathbb{E}_{\mu_a}[v(\mathcal{X}^1)]=\frac{\int v(\mathcal{X}^1)\prod_{j=1}^n\phi_2(\mathcal{X}^j,\Omega)\times\phi_{2n}\left((\tilde{X}-\tilde{\mathcal{X}}),\Sigma_a\right)\mathrm{d}\tilde{\mathcal{X}}}{\int\prod_{j=1}^n\phi_2(\mathcal{X}^j,\Omega)\times\phi_{2n}\left((\tilde{X}-\tilde{\mathcal{X}}),\Sigma_a\right)\mathrm{d}\tilde{\mathcal{X}}}$$

As $\Sigma_a\to\Sigma$, the measure $\phi_{2n}\left((\tilde{\mathcal{X}}-\tilde{X}),\Sigma_a\right)\mathrm{d}\tilde{\mathcal{X}}$ weakly converges to

$$\phi_2\left((\mathcal{X}^1-X^1),T^{-1}\right)\times\prod_{i=2}^n 1_{\mathcal{X}^i-X^i=\mathcal{X}^1-X^1}\mathrm{d}\tilde{\mathcal{X}},$$

the measure where $\epsilon^i = \epsilon^j$ for all i, j. Therefore, by weak convergence we have

$$\begin{aligned}\mathbb{E}_{\mu_a}[v(\mathcal{X}^1)] \to & \frac{\int v(\mathcal{X}^1)\prod_{j=1}^n \phi_2(\mathcal{X}^j,\Omega)\times\phi_2\left((\mathcal{X}^1 - X^1), T^{-1}\right)\times\prod_{i=2}^n 1_{\mathcal{X}^i - X^i = \mathcal{X}^1 - X^1}\mathrm{d}\tilde{\mathcal{X}}}{\int \prod_{j=1}^n \phi_2(\mathcal{X}^j,\Omega)\times\phi_2\left((\mathcal{X}^1 - X^1), T^{-1}\right)\times\prod_{i=2}^n 1_{\mathcal{X}^i - X^i = \mathcal{X}^1 - X^1}\mathrm{d}\tilde{\mathcal{X}}} \\ =& \mathbb{E}_{\mu}[v(\mathcal{X}^1)]\end{aligned}$$

as $\Sigma_a \to \Sigma$. This completes the proof.

A.5. Theorem 1 and Bordalo et al. (2013)

The model in Bordalo et al. (2013) is very general and can violate our Theorem 1. In their model, the utility of $\mathbf{x}$ out of the set $\{\mathbf{x},\mathbf{y},\mathbf{z}\}$ is

$$u(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{w}\}) = \begin{cases} 2\frac{\delta x_1^* + x_2^*}{1+\delta}, & \sigma(x_1^*, \frac{x_1^*+y_1^*+w_1^*}{3}) < \sigma(x_2^*, \frac{x_2^*+y_2^*+w_2^*}{3}); \\ x_1^* + x_2^* & \sigma(x_1^*, \frac{x_1^*+y_1^*+w_1^*}{3}) = \sigma(x_2^*, \frac{x_2^*+y_2^*+w_2^*}{3}); \\ 2\frac{x_1^* + \delta x_2^*}{1+\delta} & \sigma(x_1^*, \frac{x_1^*+y_1^*+w_1^*}{3}) > \sigma(x_2^*, \frac{x_2^*+y_2^*+w_2^*}{3}), \end{cases}$$

where $\delta \in (0,1)$ and $\sigma(a,b) = \frac{|a-b|}{a+b}$. The same formula holds for $\mathbf{y}$ and $\mathbf{w}$ by interchanging the letters. Consider the case when $\mathbf{y}^* = (12,13)$, $\mathbf{x}^* = (13,12.1)$, and $\mathbf{w}^* = (1,11)$. Then we have $\sigma(y_1^*, \frac{x_1^*+y_1^*+w_1^*}{3}) > \sigma(y_2^*, \frac{x_2^*+y_2^*+w_2^*}{3})$ and $\sigma(x_1^*, \frac{x_1^*+y_1^*+w_1^*}{3}) > \sigma(x_2^*, \frac{x_2^*+y_2^*+w_2^*}{3})$.

$$u(\mathbf{y},\{\mathbf{x},\mathbf{y},\mathbf{w}\}) = 2\frac{y_1^* + \delta y_2^*}{1+\delta} = 2\frac{12+\delta 13}{1+\delta} < 2\frac{\delta 12 + 13}{1+\delta} = 2\frac{x_1^* + \delta x_2^*}{1+\delta} = u(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{w}\}).$$

Since the utility of $\mathbf{y}$ is smallest possible, and the utility of $\mathbf{x}$ is the maximum possible, so the relative utility of $\mathbf{x}$ over $\mathbf{y}$, $u(\mathbf{x},\{\mathbf{x},\mathbf{y},\mathbf{w}\}) - u(\mathbf{y},\{\mathbf{x},\mathbf{y},\mathbf{w}\})$, is at maximum. Therefore, replacing $\mathbf{w}$ with any $\mathbf{z}$ for which $\mathbf{z}^* = \mathbf{w}^* + \lambda\Delta$ can only reduce the relative utility of $\mathbf{x}$ over $\mathbf{y}$, and hence cannot further induce the decision maker to choose $\mathbf{x}$ (when $\mathbf{z}$ and $\mathbf{w}$ are phantom options). Therefore, this model can violate our Theorem 1. And it does not predict the empirical findings in Wedell (1991) that when replacing such a $\mathbf{w}$ with $\mathbf{z}$ for which $z_1^* \approx x_1^* > y_1^*$ and $z_2^* < x_2^* < y_2^*$ will induces the decision maker to choose $\mathbf{x}$.

Acknowledgments

The author is grateful to John Nachbar, Paulo Natenzon, and Jonathan Weinstein for the inspiring discussions and guidance. The author thanks Heski Bar-Isaac, Christopher Chambers, Wiola Dziuda, Joseph Goodman, Liang Guo, Shaowei Ke, Robyn LeBoeuf, David Levine, Anqi Li, Stephen Nowlis, Brian Rogers and Marciano Siniscalchi for useful comments. The author also thanks participants at the Spring 2016 Midwest Economic Theory Meeting, 2016 Econometric Society China Meeting and 2016 SAET Conference for their comments. All errors are the author's responsibility. The author acknowledges support from the Center for Research in Economics and Strategy (CRES), in the Olin Business School, Washington University in St. Louis, and the Economic Theory Center at Washington University in St. Louis.

References

Aguiar, V. 2015. Stochastic choice and fuzzy attension. *working paper*

Ariely, D., G. Loewenstein, D. Prelec. 2003. 'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences". *The Quarterly Journal of Economics* **118**, 73-106.

Bordalo, P., N. Gennaioli, A. Shleifer. 2013. Salience and consumer Choice. *Journal of Political Economy*.

Block, H.D. and Marschak, J. 1960. Random orderings and stochastic theories of responses. in *Contributions to probability and statistics*. I. Olkin, S. Ghurye, W. Hoeffding, W. Madow and H. Mann (Eds.). pp 97-132.

Busemeyer, J.R. J.T. Townsend. 1993. Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment, *Psychological Review* **100** 432–459.

de Clippel, G. and E. Kfir. 2012. Reason-based choice: a bargaining rationale for the attraction and compromise effects *Theoretical Economics*, **7** 125-162.

Echenique, Saito, Tserenjigmid. 2015. The Perception Adjusted Luce Model. *working paper*.

Falmagne, J.C. 1978. A Representation Theorem for Finite Random Scale System. it Journal of Mathematical Psychology, **18**, 52-72

Fehr, E., and A. Rangel. 2011. Neuroeconomic Foundations of Economic Choice–Recent Advances. *Journal of Economic Perspectives*, **25** 3-30.

Freeman, D. J. 2017. Preferred personal equilibrium and simple choices. *Journal of Economic Behavior & Organization*, **143** 165-172.

Fudenberg, D., P. Strack, T. Strzalecki. 2015. Stochastic choice and optimal sequential sampling. *working paper*.

Green, P. E., and V. Srinivasan. 1978. Conjoint Analysis in Consumer Research: Issues and Outlook. *Journal of Consumer Research* Vol. 5, No. 2, pp. 103-123.

Gul, F., P. Natenzon and W. Pesendorfer. 2014. Random Choice as Behavioral Optimization. *Econometrica* **82** 1873-1912.

Guo, L. 2016. Contextual Deliberation and Preference Construction. *Management Science* **62** 2977-2993

Hausman, J.A. D.A. Wise. 1978. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica*

Hedgcock, W., A.R. Rao and H.A. Chen. 2009. Could Ralph Nader's Entrance and Exit Have Helped Al Gore? The Impact of Decoy Dynamics on Consumer Choice. *Journal of Marketing Research*

Highhouse, S. 1996. Context-dependent selection: The effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes* **65** 68-76.

Hsee, C. K. 1996. The evaluability hypothesis: An explanation of preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes* **67** 247-257.

Hsee, C. K., G. F. Loewenstein, S. Blount and M. H. Bazerman. 1999. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin* **125** 576-590.

Huber, J., J. W. Payne, C. Puto. 1982. Adding asymmetrically dominated alternatives: Violation of regularity and similarity hypothesis. *Journal of Consumer Research*

Iyengar S.S., Lepper M.R. (2000) When choice is demotivating: Can one desire too much of a good thing? *J. Personality Soc. Psych.* 79 (6): 995-1006.

Kamenica, E. 2008. Contextual inference in markets: On the informational content of product lines. *American Economic Review* **98** 2127-2149

Koszegi, B., Rabin, M. 2006. A model of reference-dependent preferences. *Quarterly Journal of Economics*

Koszegi, B., Szeidl, A. 2013. A Model of Focusing in Economic Choice. *Quarterly Journal of Economics* 53-104

Leland, J. W. 1994. Generalized Similarity Judgements: An Alternative Explanation for Choice Anomalies. *Journal of Risk and Unertainty* 9: 151-172.

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

Manzini, P., Mariotti, M. 2007. Sequentially Rationalizable Choice. *American Economic Review* **97**. No. 5

Masatlioglu, Y., D. Nakajima, E.Y. Ozbay. 2012 Revealed attention. *The American Economic Review* **102** 2183–2205.

Natenzon, P. 2019. Random Choice and Learning. *Journal of Political Economy* **127** 419-457.

Ok E. A., P. Ortoleva, G. Riella. 2015. Revealed (P)Reference Theory. *AMERICAN ECONOMIC REVIEW* **105** 299-321.

Pettibone, J. C., D. H. Wedell. 2000. Examining Models of Nondominated Decoy Effects across Judgment and Choice. *Organizational Behavior and Human Decision Processes* **81** No.2 300-328.

Pettibone, J. C., D. H. Wedell. 2007. Testing Alternative Explanations of Phantom Decoy Effects. *Journal of Behavioral Decision Making* **20** 323–341

Plous, Scott (1993). *The Psychology of Judgment and Decision Making*. McGraw-Hill. 38-41.

Pratkanis, A. R., P. H. Farquhar. 1992. A brief history of research on phantom alternatives: Evidence for several empirical generalization about phantoms. *Basic and applied social psychology* **13** 103-122

Ratcliff, R. 1978. A Theory of Memory Retrieval. *Psychological Review* **85** 59–108.

Rieskamp, J., J. R. Busemeyer, B. A. Mellers. 2006. Extending the bounds of rationality: evidence and Theories of preferential choice *Journal of Economic Literature*

Ravid, D. 2015. Focus, Then Compare *working paper*.

Roe, Robert M., Jerome R. Busemeyer, and James T. Townsend. 2001. Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making. *Psychological Review*, 108(2): 370-92.

Schwarz, N., and Bless, H. (1992). *Scandals and the Public's Trust in Politicians: Assimilation and Contrast Effects*. Personality and Social Psychology Bulletin, **18**, 574–579.

Simonson, I. 1989. Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research* **16** 158-174

Sivakumar, K., J. Cherian. 1995. Role of product entry and exit on the attraction effect. *Marketing Letters* **6**, 45-51.

Soltani, A., B. De Martino, C. Camerer. 2012. A RangeNormalization Model of Context-Dependent Choice: A New Model and Evidence. *PLoS computational biology* **8**

Tserenjigmid, G. 2019. Choosing with the Worst in Mind: A Reference-Dependent Model. *Journal of Economic Behavior & Organization* **157**, 631 - 652

Thurstone, L.L., 1927. Psychophysical analysis. *The American Journal of Psychology* **38** 368-389.

Tversky, A. 1969. Intransitivity of Preferences. *Psychological Review* **76** (1), 31-48.

Tversky, A. 1972. Elimination by Aspects: A Theory of Choice. *Psychological Review* **79** 281-299.

Tversky, A., J.E. Russo. 1969. Substitutability and similarity in binary choices. *Journal of mathematical Psychology* 1-12.

Tversky, A., I. Simonson. 1993. Context-Dependent Preferences *Management Science* **39** (10) 1179-1189.

Wedell, D. H. 1991. Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **17** (4) 767-778.

Wernerfelt, B. 1995. A rational reconstruction of the compromise effect: Using market data to infer utilities. *Journal of Consumer Research* 627–633.

Woodford, M. 2014. Stochastic choice: An optimizing neuroeconomic model *The American Economic Review* **104** 495–500.

Usher, M., J. L. McClelland. 2004. Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review* **111**