# A rational account of the repulsion effect

Rahul Bhui[1,*] and Yang Xiang[1,2]

[1]Sloan School of Management, MIT

[2]Department of Psychology, Harvard University

[*]Corresponding author: rbhui@mit.edu

November 9, 2022

Author note: We gratefully acknowledge funding from the MIT Sloan School of Management and the Office of Naval Research.

Abstract

The repulsion effect occurs when the presence of an inferior option (the decoy) decreases the attractiveness of the option that dominates it (the target), in puzzling contrast to the classic attraction effect. In this paper, we formally develop and experimentally test a normative account of the repulsion effect. Our theory is based on the idea that the true values of options are uncertain and must be inferred from available information, which includes the properties of other options. A low-value decoy can signal that the target also has low value when both are believed to be generated by a similar process. We formalize this logic using a hierarchical Bayesian cognitive model that makes predictions about how the strength of the repulsion effect should vary with statistical properties of the decision problem. This theory can help account for several documented phenomena linked to the repulsion effect, as well as new experimental data. Our results illuminate key drivers of context dependence across both economic and perceptual judgment and sharpen our understanding of when decoys can be detrimental.

*Keywords:* repulsion effect, context dependence, Bayesian, hierarchical, judgment and decision making

## A rational account of the repulsion effect

The exact same person can face the exact same stimulus but feel completely differently about it depending on the context. When people make decisions, such context dependence can lead to surprising preference reversals. For instance, adding a clearly inferior item (the "decoy") to the choice set can make the option which dominates it (the "target") more appealing; this phenomenon is known as the *attraction effect* (Huber, Payne, & Puto, 1982). The attraction effect has remained front and center in the study of decision making for decades. It has helped to identify the empirical boundaries of rational choice theory (Huber, Payne, & Puto, 2014) and motivated the multidisciplinary search for deeper principles of context-sensitive judgment across marketing, psychology, economics, neuroscience, and biology (e.g., Bhui, Lai, & Gershman, 2021; Bordalo, Gennaioli, & Shleifer, 2013; Hedgcock & Rao, 2009; Lea & Ryan, 2015; V. Li, Michael, Balaguer, Castañón, & Summerfield, 2018; Roe, Busemeyer, & Townsend, 2001; Simonson, 1989; Spektor, Bhatia, & Gluth, 2021; Tversky & Simonson, 1993).

However, the robustness of the attraction effect has been seriously questioned by recent work. Frederick, Lee, and Baskin (2014) found a broad lack of evidence for the effect in 38 studies, including replications of several influential paradigms. Yang and Lynn (2014) similarly found little reliable evidence of the effect across 91 attempts to produce it. Trendl, Stewart, and Mullett (2021) found a precisely estimated null effect in an experiment designed to meet the ideal test conditions laid out by Huber et al. (2014). This research suggests the attraction effect may not be as common as popularly thought.

Most strikingly, the opposite pattern has been observed, in which the inferior decoy makes the dominating target seem worse—a *repulsion effect* (Frederick et al., 2014). The repulsion effect has been documented in preferential choices between consumer goods (Frederick et al., 2014; Liao, Chen, Lin, & Mo, 2020), reinforcement learning paradigms with abstract options (Ert & Lejarraga, 2018; Spektor, Gluth, Fontanesi, & Rieskamp, 2019), perceptual decisions like picking the largest rectangle (Spektor, Kellen, & Hotaling,

2018), and even hybrid tasks with preferential decisions presented using perceptual stimuli (Brendl, Atasoy, & Samson, in press; Spektor, Kellen, & Klauer, 2022). These anomalous yet wide-ranging findings indicate the need for alternative mechanisms of context dependence.[1] Attempts to model the repulsion effect so far have been confined to mechanistic approaches that trace out the deliberation process behind decision making and have met with limited success (Spektor et al., 2021, 2022).

Why does the repulsion effect occur? Some informally propose that inferior decoys may "contaminate" other options with similar attributes. This idea is referred to as the *tainting hypothesis* (Frederick et al., 2014; Simonson, 2014). However, despite its intuitive appeal, it is not clear why exactly such a mechanism should exist or what factors should modulate its strength.

In this paper, we formally develop and experimentally test a normative account of the repulsion effect which can be viewed as a Bayesian formulation of the tainting hypothesis. Our theory is based on the idea that the true values of options are uncertain and must be inferred from available information, which includes the properties of other options. A low-value decoy can signal that the target also has low value when both are believed to be generated by a similar process. For example, a bad product can signal bad brand quality, indicating that other products from the same brand are probably not as good as they appear. People may likewise draw shared inferences about two restaurants from the same franchise, two job candidates from the same training program, two movies from the same studio, or two fruits from the same basket. We capture this logic in a hierarchical Bayesian cognitive model (Tenenbaum, Griffiths, & Kemp, 2006). Because uncertainty can arise

---

[1] The repulsion effect should not be conflated with the similarity effect (Tversky, 1972) even though both describe a relative decrease in preference for the target option similar to the decoy. The key difference between the two effects is that the decoy is inferior to the target in the former case but of comparable value in the latter. Like Spektor et al. (2019), we saw no support for a similarity-effect interpretation because participants preferred the decoy substantially less than the target, and were thus evidently able to distinguish the two.

from sources as diverse as limited information, sensory noise, and constraints on cognitive processing, this mechanism applies to both economic and perceptual judgment (Woodford, 2020). Even when groups are not explicitly known, people may form inferences about which items were likely generated by the same underlying statistical process (Anderson, 1991). In this perspective, the repulsion effect may reflect an adaptive function which improves decision quality, because the context provides information that helps to interpret the available options (McKenzie, Sher, Leong, & Müller-Trede, 2018).

Our theory is consistent with an array of research from across the cognitive and behavioral sciences. Several empirical studies in quantitative marketing use related models to analyze how consumers learn about different products from the same brand or category (Ching, Erdem, & Keane, 2013; Erdem, 1998; Erdem & Chang, 2012; Sridhar, Bezawada, & Trivedi, 2012). More broadly, an expanding body of work in cognitive science fruitfully applies hierarchical Bayesian models[2] to a wide range of domains such as reinforcement learning (Acuña & Schrater, 2010), sensory learning (Mathys, Daunizeau, Friston, & Stephan, 2011; Mathys et al., 2014), advice taking (Diaconescu et al., 2014), motion perception (Bill, Pailian, Gershman, & Drugowitsch, 2020), linguistics (Xu & Tenenbaum, 2007), and beyond (e.g., Griffiths & Tenenbaum, 2009; Sharp, Fradkin, & Eldar, 2022; Tenenbaum, Kemp, Griffiths, & Goodman, 2011).

The main contribution of our paper is to rigorously develop a normative account of the repulsion effect. By doing so, we provide a unified explanation for its existence across

---

[2] Our hierarchical cognitive modeling approach should not be confused with either hierarchical Bayesian statistical modeling (Gelman et al., 2013) or hierarchical decision making processes (Evangelidis, Levav, & Simonson, 2018; Tversky, Sattath, & Slovic, 1988). Purely statistical hierarchical models are only meant to capture individual heterogeneity for purposes of improved parameter estimation. They make no direct claims about cognitive principles or processes, even though the mathematical structures we use are the same. The process models trace out a sequence of broad decision strategies taken by an agent to arrive at a choice. They are not specifically about inferential principles, though these could be invoked by the strategies they contain.

economic and perceptual domains, as well as transparently expose the conditions under which it should emerge. Our formulation also enables a sharper specification of what similarity entails, and allows this *Bayesian tainting hypothesis* to be precisely laid out in a variety of environments. We maximize transparency by deriving closed-form analytical expressions that encapsulate the repulsion effect.

In what follows, we present our theory and show how it can help account for a number of previous findings, such as the emergence of the repulsion effect in decisions from experience (Ert & Lejarraga, 2018; Spektor et al., 2019), the double decoy effect (Daviet & Webb, 2020), the phantom decoy effect (e.g., Pettibone & Wedell, 2007; Pratkanis & Farquhar, 1992), and the non-monotonic impact of decoy distance on the repulsion effect (Liao et al., 2020). We also conduct four new experiments using consumer choice tasks and value estimation paradigms which demonstrate that the repulsion effect is stronger when the target and decoy are believed to come from the same group and have more correlated values, as predicted by the theory. Finally, we discuss managerial implications.

Our work yields several benefits. First, our theory helps to account for empirical variation in the effect observed across past studies and new experiments in a unified way. Second, it enables us to specify a version of the tainting hypothesis in any setting where an appropriate causal model can be written down. Third, it offers a normative rationale for the repulsion effect, linking it with branches of research that develop probabilistic accounts of cognition (e.g., Doya, Ishii, Pouget, & Rao, 2007; Griffiths, Kemp, & Tenenbaum, 2008; Oaksford & Chater, 2007) and seek adaptive explanations for biases in judgment (e.g., Bhui et al., 2021; Gershman, Horvitz, & Tenenbaum, 2015; Griffiths, Lieder, & Goodman, 2015; Lewis, Howes, & Singh, 2014; Lieder & Griffiths, 2020; Summerfield & Parpart, 2022).

## Bayesian Models of Context Effects

Bayesian principles have been used before to explain contextual preference reversals like the attraction effect. For instance, when decision makers are uncertain about how

valuable each attribute is, the presence of a decoy can indicate that the target offers better than fair market value (Ahmad & Yu, 2015; Shenoy & Yu, 2013; see also Wernerfelt, 1995, Kamenica 2008, S. Li and Yu 2018, and Sher and McKenzie 2014). Alternatively, because the decoy is more similar to the target than to the competitor, it may be easier to compare (Markman & Medin, 1995; Tversky & Russo, 1969), thus providing ordinal information which raises the probability that the target is the best option (Howes, Warren, Farmer, El-Deredy, & Lewis, 2016; Natenzon, 2019). Bayesian inference may also be combined with the non-Bayesian assumption that preferences reflect deviations from one's expectations (Rigoli, Mathys, Friston, & Dolan, 2017). However, while all of these models appeal to inferential mechanisms not unlike our own, they are designed to generate attraction effects and do not naturally produce repulsion effects.[3]

Our approach is closest to Bordley's (1992) analysis of context sensitivity in the domain of risky choice (see also Luce and Raiffa 1957, p. 288). Following Viscusi (1989), he supposes that people do not take stated lottery outcomes at face value, and instead form assessments by adjusting their prior beliefs about lotteries toward the stated outcomes in a Bayesian fashion. He considers how various anomalies including one like the repulsion effect can occur when some lotteries have correlated outcomes.[4] For example, suppose a customer at a restaurant must choose between beef, fish, and chicken, and that the quality of only the latter two depend heavily on the competence of the chef. Beef might be preferred to fish when they are the only two options; but if chicken were also available, the presence of both fish and chicken would indicate that the chef is competent since the restaurant would not likely advertise two bad dishes. In that case, the customer's assessment of the fish would improve, leading them to select fish over beef from a menu of

[3] The mechanism we focus on conversely produces a repulsion effect but not an attraction effect, though in future work these different components may be integrated as they are compatible and not mutually exclusive.

[4] This connection has gone unnoticed in the literature on repulsion effects to date.

all three meals even when they would select beef over fish in the absence of chicken. As in our account of the repulsion effect, the presence of a third item serves as a signal of another item's value, leading to a preference reversal that violates independence of irrelevant alternatives (Luce, 1959). Our formulation would encode correlated outcomes via the hierarchical structure, where the group mean denotes the chef's competence defined as the average quality of the meals they serve. In addition to the many new applications we present here, we move beyond Bordley (1992) from a technical standpoint in multiple ways. We allow options to consist of multiple attributes and we derive similarity itself within the model. These advances are important for capturing observed properties of repulsion effects as described below.

## A Bayesian Model of the Repulsion Effect

We lay out our model in a two-attribute setting where the attributes are continuous variables such as numeric ratings of quality or price and the objective function takes on an additive form. Due to the flexibility of the Bayesian framework, our theory can be applied to a much wider set of attribute types and objective functions through appropriate modification of the generative process. We will briefly explore such modifications later in the subsection on Model Extensions.

Consider an agent evaluating a target option and a competitor in the presence of a decoy; call these $T$, $C$, and $D$ respectively. Each option $i \in \{T, C, D\}$ is described by two observable attributes, which we denote by $x_{i,1}$ and $x_{i,2}$ (and collectively by $x_i$). In traditional models of multi-attribute choice with additive utility, the value of an option would be given by $v_i = \omega_1 x_{i,1} + \omega_2 x_{i,2}$ for some coefficients $\omega_1$ and $\omega_2$ that reflect the subjective importance placed on each attribute. In the standard view, agents place intrinsic value on various attributes. Hence, customers might be thought to place a particular intrinsic value on a restaurant with a high quality rating. We propose instead that the observable attributes are merely superficial manifestations of an option's latent properties,

and the agent does not have perfect access to the true values of options. In this view, a high quality rating is merely a signal of how good the customer's experience would be. It is noisy, since a high rating does not guarantee a great experience, but it is informative, and helps the customer construct an estimate of value. We take the agent's evaluation to be the expected value of each option conditional on the set of options (defined in terms of their observable attributes), $E[v_i \mid x_j \forall j]$.

We assume the agent forms evaluations by inverting a hierarchical Bayesian generative model, depicted in Figure 1. In this model, options are described by two latent properties as illustrated in Figure 2—value, $v_i \in \mathbb{R}$, and what we call "style," $s_i \in \mathbb{R}$. We define style as a value-orthogonal characteristic that determines the relative balance between attributes 1 and 2, holding value fixed. For example, two restaurants may be equally desirable overall even though one has better service if it is exactly offset by worse food; in our terminology, they have the same value but different styles. Technically speaking, value indicates which indifference curve an option resides on, while style indicates an option's position along that indifference curve. This kind of decomposition has been used before under other names (e.g., Berkowitsch, Scheibehenne, Rieskamp, & Matthäus, 2015; Rooderkerk, Van Heerde, & Bijmolt, 2011; Wedell, 1991). Although style does not bear directly on value, it will do so indirectly by affecting how similar the target and decoy are judged to be in the absence of explicit labels (see the subsection on Latent Group Inference).

An option's observable attributes are generated from its latent properties according to a common objective function and corrupted by noise. Specifically, we model this observation noise in two steps, roughly following Shenoy and Yu (2013) and Ahmad and Yu (2015): we first add Gaussian noise to the option's latent properties of style and value, and then map these noisy variables to attribute space by inverting the objective function. Accordingly, let $\hat{v}_i \equiv v_i + \varepsilon_{i,v}$ and $\hat{s}_i \equiv s_i + \varepsilon_{i,s}$ where $\varepsilon_{i,v} \sim \mathcal{N}(0, \sigma_v^2)$ and $\varepsilon_{i,s} \sim \mathcal{N}(0, \sigma_s^2)$ are independent noise terms. With an additive objective function, we thus construct attribute

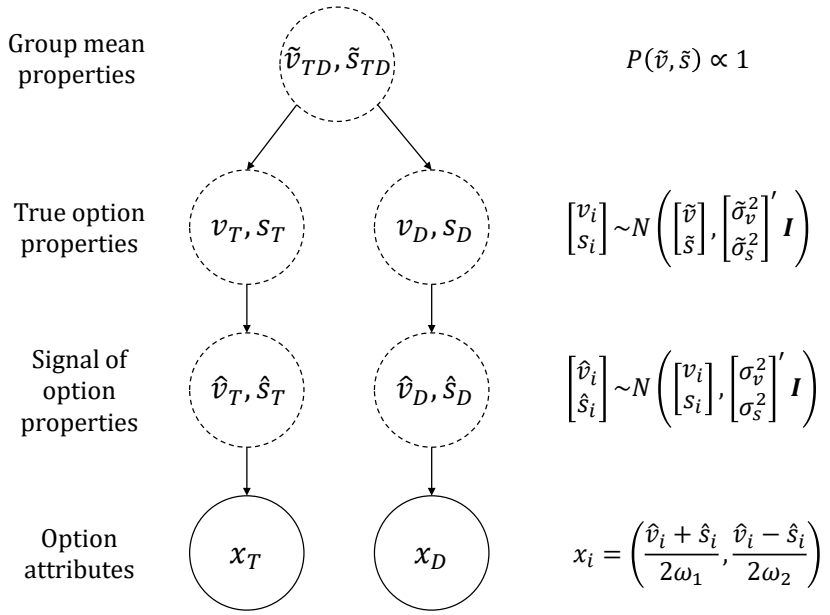


*Figure 1*. Hierarchical Bayesian generative model leading to the repulsion effect. Solid lines denote observed variables and dashed lines denote latent variables. Note that we depict the signals ($\hat{v}$, $\hat{s}$) as latent for the sake of generality, even though in our setup they are effectively observed due to a one-to-one mapping with the attributes ($x$).

magnitudes such that $\hat{v}_i = \omega_1 x_{i,1} + \omega_2 x_{i,2}$ and $\hat{s}_i = \omega_1 x_{i,1} - \omega_2 x_{i,2}$. Inverting these expressions yields $x_{i,1} = \frac{\hat{v}_i + \hat{s}_i}{2\omega_1}$ and $x_{i,2} = \frac{\hat{v}_i - \hat{s}_i}{2\omega_2}$. Thus, the attribute magnitudes are increasing in underlying value and differentiated from each other based on style. The denominator stems from our assumption that noise occurs in the latent space, and so the same amount of noise translates into a small difference for attributes that are more important and a large difference for attributes that are less important. We make the assumption that noise occurs in the value-style space for mathematical simplicity, but could suppose that noise directly impacts the attribute level[5] without greatly altering the results. Our construction implies that $(x_{i,1}, x_{i,2})$ and $(\hat{v}_i, \hat{s}_i)$ are isomorphic, and that $\hat{v}_i$ and

[5] Noise in the attributes would entail that $x_{i,1} = \frac{v_i + s_i}{2\omega_1} + \varepsilon_{i,1}$ and $x_{i,2} = \frac{v_i - s_i}{2\omega_2} + \varepsilon_{i,2}$ where $\varepsilon_{i,1} \sim \mathcal{N}(0, \sigma_1^2)$ and $\varepsilon_{i,2} \sim \mathcal{N}(0, \sigma_2^2)$. Both kinds of noise could even be included simultaneously, at some expense of clarity.

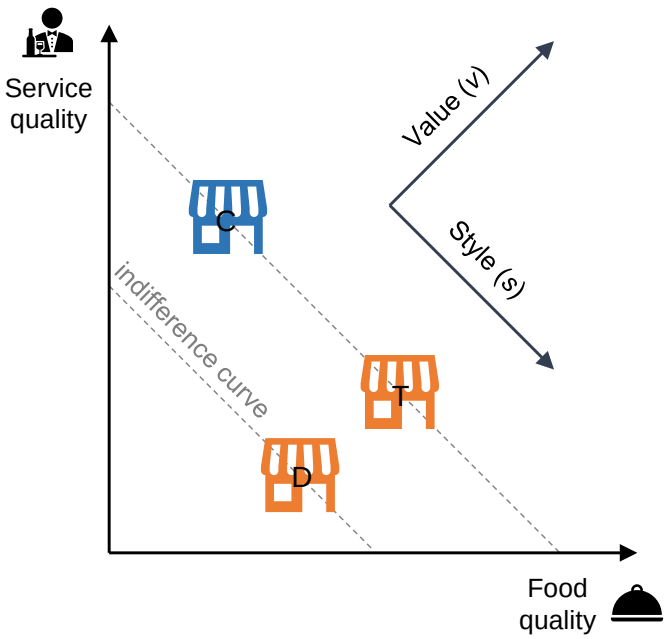


*Figure 2*. Example illustration of relevant dimensions, with restaurant options characterized by the attributes of food quality and service quality. Value and style dimensions are orthogonal to each other, with the style axis running parallel to the indifference curves. T = target, D = decoy, C = competitor. Color indicates the grouping of the restaurants based on similarity in style.

$\hat{s}_i$ are independent, so

$$\begin{bmatrix} \hat{v}_i \\ \hat{s}_i \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} v_i \\ s_i \end{bmatrix}, \begin{bmatrix} \sigma_v^2 \\ \sigma_s^2 \end{bmatrix}' \mathbf{I} \right) \tag{1}$$

where $\mathbf{I}$ is the identity matrix. We can thus work directly with $\hat{v}_i$ and $\hat{s}_i$.

We further suppose that the latent properties are themselves drawn from higher-level prior distributions:

$$\begin{bmatrix} v_i \\ s_i \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \tilde{v}_i \\ \tilde{s}_i \end{bmatrix}, \begin{bmatrix} \tilde{\sigma}_v^2 \\ \tilde{\sigma}_s^2 \end{bmatrix}' \mathbf{I} \right). \tag{2}$$

In a consumer choice setting, for instance, $\tilde{v}_i$ might reflect brand quality—the average value of a random product from the brand that option $i$ belongs to. Then $\tilde{\sigma}_v$ would capture the

degree to which different products from the brand vary in value, and $\sigma_v$ would capture the degree of noise in how precisely the observable attributes of a product signal its latent value. Importantly, the group-level means $\tilde{v}_i$ and $\tilde{s}_i$ are themselves unknown and must be inferred. For simplicity, we suppose $\tilde{\sigma}_v$ and $\tilde{\sigma}_s$ are known and common to all groups, and assume noninformative uniform hyperpriors for $\tilde{v}_i$ and $\tilde{s}_i$ given $\tilde{\sigma}_v$ and $\tilde{\sigma}_s$ respectively.

A key assumption of our theory is that people believe the target and decoy are generated by a shared underlying process, which implies they share the same group-level means, $\tilde{v}_T = \tilde{v}_D = \tilde{v}_{TD}$ and $\tilde{s}_T = \tilde{s}_D = \tilde{s}_{TD}$. As a consequence, the attributes of the decoy provide information about the value of the target. We suppose for the moment that explicit labels are provided which induce this belief, but will later relax this in the subsection on Latent Group Inference.

The agent's posterior belief about the target's underlying value, given the target and decoy attributes, is $P(v_T|x_T, x_D)$. This can be obtained by calculating the marginal distribution of $P(v_T|x_T, x_D, \tilde{v}_T)$ which is known to have a closed form solution under the above generative model (Berger, 1985, Section 4.6; Gelman et al., 2013, Section 5.4), with expected value

$$E[v_T \mid x_T, x_D, \tilde{v}_T] = \frac{\tilde{\sigma}_v^2 \hat{v}_T + \sigma_v^2 \tilde{v}_T}{\sigma_v^2 + \tilde{\sigma}_v^2}. \tag{3}$$

Marginalizing over $\tilde{v}_T$ yields

$$E[v_T \mid x_T, x_D] = \frac{\tilde{\sigma}_v^2 \hat{v}_T + \sigma_v^2 E[\tilde{v}_T \mid x_T, x_D]}{\sigma_v^2 + \tilde{\sigma}_v^2} \tag{4}$$

where

$$E[\tilde{v}_T \mid x_T, x_D] = \frac{\hat{v}_T + \hat{v}_D}{2}. \tag{5}$$

The posterior mean of target value, $E[v_T|x_T, x_D]$, is a precision-weighted average of (i) the value signal derived from the option's attributes and (ii) the posterior mean of the group value, which is the midpoint of the target and decoy value signals. That is, it is a convex

combination of the individual item expectation and the pooled group expectation. Algebra reveals that

$$E[v_T \mid x_T, x_D] = \hat{v}_T - \left(\frac{\sigma_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2}\right)\left(\frac{\hat{v}_T - \hat{v}_D}{2}\right). \tag{6}$$

The repulsion effect can be characterized by comparing this quantity with the posterior mean of $v_T$ when the decoy is not present, $E[v_T|x_T]$, which is simply $\hat{v}_T$:

$$E[v_T \mid x_T, x_D] - E[v_T \mid x_T] = -\left(\frac{\sigma_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2}\right)\left(\frac{\hat{v}_T - \hat{v}_D}{2}\right). \tag{7}$$

The decoy is inferior to the target, meaning $\hat{v}_T > \hat{v}_D$, and so this expression is negative. Hence, $E[v_T \mid x_T, x_D] < E[v_T \mid x_T]$: the decoy exerts a negative influence on the perceived value of the target.[6] The result is visualized in Figure 3.

This expression not only encodes a repulsion effect, but also transparently reveals the conditions necessary for it to emerge. The effect depends crucially on the cohesiveness of the group, reflected in the parameter $\tilde{\sigma}_v$, and the level of uncertainty about option values, reflected in $\sigma_v$. The decoy and target values must be informatively drawn from the same group, meaning $\tilde{\sigma}_v$ must be small; the effect vanishes as $\tilde{\sigma}_v$ grows large. Furthermore, the underlying values must be uncertain enough for information to matter, meaning $\sigma_v$ must be relatively large; the effect vanishes as $\sigma_v$ becomes small. Note also that the downward bias in target evaluation is mirrored by an upward bias in decoy evaluation, as the symmetric derivation (assuming items have the same noise levels) entails that $E[v_T \mid x_T, x_D] - E[v_T \mid x_T] = -(E[v_D \mid x_D, x_T] - E[v_D \mid x_D])$.

### Latent Group Inference

The above expressions assume the decision maker knows that the target and decoy come from the same group. For instance, when describing bottled water options, Frederick

[6] The posterior belief when the decoy is present is also more precise. Greater precision could increase utility for risk averse agents, as Roberts and Urban (1988) and Erdem (1998) point out in the context of product branding. While this may be a meaningful factor especially in economic decisions, it would oppose the reduction in mean value we focus on above, and thus cannot produce a repulsion effect.

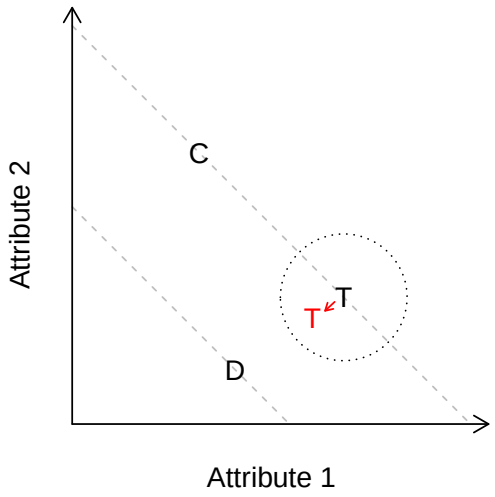


*Figure 3*. Illustration of the Bayesian repulsion effect. T = target, D = decoy, C = competitor. The dashed lines represent indifference curves, the dotted circle represents uncertainty about the target's latent properties, and the red T represents the posterior mean of the target's properties. Note the mechanism is depicted in attribute space for illustration, though it may occur in the latent value-style space.

et al. (2014) explicitly label both the target and decoy as kinds of spring water, in contrast to the competitor; or when depicting microwave popcorn, the target and decoy are shown to be from the same brand which is different from the competitor. However, such information is not always given. For example, the rectangles in Spektor et al.'s (2018) perceptual task are not overtly assigned to particular groups. Even a label might be considered as any other attribute, just one that strongly predicts group identity. The decision maker must ultimately group items together of their own accord based on their sense of similarity.

We cast these similarity assessments as Bayesian inference of latent group structure (Austerweil, Gershman, Tenenbaum, & Griffiths, 2015). Models in this vein have been fruitfully applied to understand judgment in various settings such as category assignment (Anderson, 1991), social evaluation (Gershman, Pouncy, & Gweon, 2017; Lau, Gershman, & Cikara, 2020; Lau, Pouncy, Gershman, & Cikara, 2018), multisensory perception (Cao, Summerfield, Park, Giordano, & Kayser, 2019; Körding et al., 2007; Sato, Toyoizumi, &

Aihara, 2007; Shams & Beierholm, 2010), reinforcement learning (Gershman, Blei, & Niv, 2010; Gershman, Norman, & Niv, 2015), and anchoring (Wilson, Arora, Zhang, & Griffiths, 2021). Our approach is also inspired by that of Kemp, Bernstein, and Tenenbaum (2005) who show how several common similarity measures can be derived from a unified Bayesian framework. Rather than being an alternative to classic conceptions of similarity (Bhui, 2018; Tversky, 1977), Bayesian principles can provide deeper foundations that inform us about the properties of similarity functions.

To incorporate uncertainty about group membership, we apply the law of total expectation to the posterior mean of target value, splitting it into cases where the target and decoy come from the same group (i.e., the same higher-level distribution) versus different groups (see Figure 4). To simplify, we assume the competitor is so dissimilar from the others as to be almost definitely from a different group, though this can be relaxed naturally by explicitly considering all possible combinations of options (see subsection on Model Extensions). Notice that the dimension of style is important here so that the target and the decoy are classified together on the basis of their stylistic similarity, rather than the target and the competitor which are typically closer in terms of value. Let $g$ represent the number of groups, equaling 1 when the target and decoy are from the same group and 2 when they are from different groups. Then

$$
\begin{aligned}
E[v_T|x_T, x_D] &= E[v_T|x_T, x_D, g = 1]P(g = 1|x_T, x_D) && (8)\\
&\quad + E[v_T|x_T, x_D, g = 2]P(g = 2|x_T, x_D) && (9)\\
&= \left[\hat{v}_T - \left(\frac{\sigma_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2}\right)\left(\frac{\hat{v}_T - \hat{v}_D}{2}\right)\right] P(g = 1|x_T, x_D) && (10)\\
&\quad + \hat{v}_T(1 - P(g = 1|x_T, x_D)) && (11)\\
&= \hat{v}_T - \left(\frac{\sigma_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2}\right)\left(\frac{\hat{v}_T - \hat{v}_D}{2}\right) P(g = 1|x_T, x_D). && (12)
\end{aligned}
$$

The resulting expression resembles equation (6), with only the emergence of a multiplier on the repulsion term scaling it by the probability that the target and decoy are thought to be from the same group. This probability can be computed from Bayes' rule,

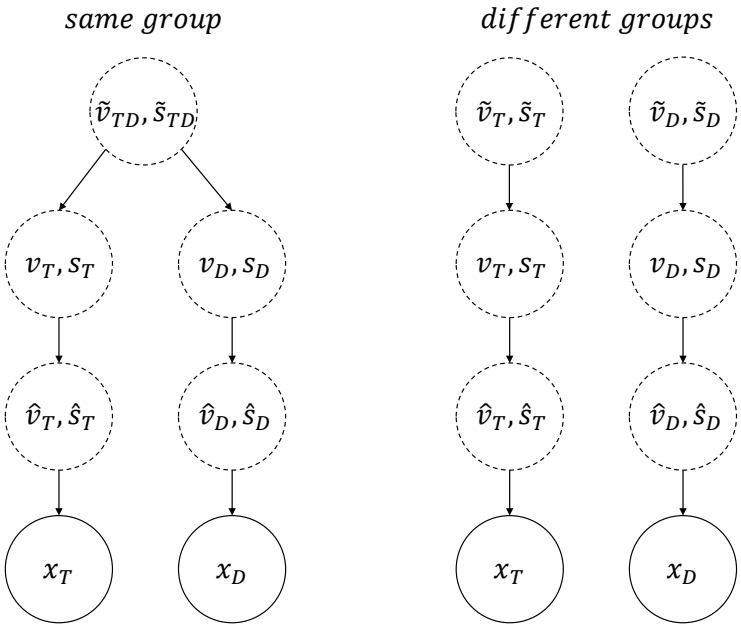


*Figure 4*. Bayesian generative models for latent group inference, when the target and decoy come from the same group (left) or different groups (right).

considering the relative probabilities that the two items came from the same generative process or two different ones. To yield an analytically tractable result, like Kemp et al. (2005) we approximate the likelihoods with the maximum *a posteriori* (MAP) values of the group-level means:

$$P(g=1|x_T,x_D) = \frac{P(x_T,x_D|g=1)P(g=1)}{P(x_T,x_D|g=1)P(g=1)+P(x_T,x_D|g=2)P(g=2)} \tag{13}$$

$$= \frac{1}{1+\dfrac{P(x_T,x_D|g=2)P(g=2)}{P(x_T,x_D|g=1)P(g=1)}} \tag{14}$$

$$\approx \frac{1}{1+\dfrac{P(x_T|\tilde{v}_T^*,\tilde{s}_T^*)P(x_D|\tilde{v}_D^*,\tilde{s}_D^*)P(g=2)}{P(x_T|\tilde{v}_{TD}^*,\tilde{s}_{TD}^*)P(x_D|\tilde{v}_{TD}^*,\tilde{s}_{TD}^*)P(g=1)}} \tag{15}$$

where the MAP estimates of the separate and shared group means are

$$(\tilde{v}_i^*,\tilde{s}_i^*) \equiv \text{argmax}_{\tilde{v}_i,\tilde{s}_i} P(x_i|\tilde{v}_i,\tilde{s}_i) = (\hat{v}_i,\hat{s}_i) \tag{16}$$

$$(\tilde{v}_{TD}^*,\tilde{s}_{TD}^*) \equiv \text{argmax}_{\tilde{v}_{TD},\tilde{s}_{TD}} P(x_T,x_D|\tilde{v}_{TD},\tilde{s}_{TD}) = \left(\frac{\hat{v}_T+\hat{v}_D}{2},\frac{\hat{s}_T+\hat{s}_D}{2}\right). \tag{17}$$

Since $\hat{v}_i$ and $\hat{s}_i$ are independently normally distributed, which entails

$$P(x_i|\tilde{v}^*, \tilde{s}^*) = \frac{1}{2\pi\sqrt{(\sigma_v^2 + \tilde{\sigma}_v^2)(\sigma_s^2 + \tilde{\sigma}_s^2)}} \exp\left(-\frac{1}{2}\left[\frac{(\hat{v}_i - \tilde{v}^*)^2}{\sigma_v^2 + \tilde{\sigma}_v^2} + \frac{(\hat{s}_i - \tilde{s}^*)^2}{\sigma_s^2 + \tilde{\sigma}_s^2}\right]\right), \tag{18}$$

we have that

$$P(x_i|\tilde{v}_i^*, \tilde{s}_i^*) = \frac{1}{2\pi\sqrt{(\sigma_v^2 + \tilde{\sigma}_v^2)(\sigma_s^2 + \tilde{\sigma}_s^2)}} \exp\left(-\frac{1}{2}\left[\frac{(\hat{v}_i - \hat{v}_i)^2}{\sigma_v^2 + \tilde{\sigma}_v^2} + \frac{(\hat{s}_i - \hat{s}_i)^2}{\sigma_s^2 + \tilde{\sigma}_s^2}\right]\right) \tag{19}$$

$$= \frac{1}{2\pi\sqrt{(\sigma_v^2 + \tilde{\sigma}_v^2)(\sigma_s^2 + \tilde{\sigma}_s^2)}}, \tag{20}$$

and for $i \in \{T, D\}$,

$$P(x_i|\tilde{v}_{TD}^*, \tilde{s}_{TD}^*) = \frac{1}{2\pi\sqrt{(\sigma_v^2 + \tilde{\sigma}_v^2)(\sigma_s^2 + \tilde{\sigma}_s^2)}} \exp\left(-\frac{1}{2}\left[\frac{\left(\hat{v}_i - \frac{\hat{v}_T + \hat{v}_D}{2}\right)^2}{\sigma_v^2 + \tilde{\sigma}_v^2} + \frac{\left(\hat{s}_i - \frac{\hat{s}_T + \hat{s}_D}{2}\right)^2}{\sigma_s^2 + \tilde{\sigma}_s^2}\right]\right) \tag{21}$$

$$= \frac{1}{2\pi\sqrt{(\sigma_v^2 + \tilde{\sigma}_v^2)(\sigma_s^2 + \tilde{\sigma}_s^2)}} \exp\left(-\frac{1}{8}\left[\frac{(\hat{v}_T - \hat{v}_D)^2}{\sigma_v^2 + \tilde{\sigma}_v^2} + \frac{(\hat{s}_T - \hat{s}_D)^2}{\sigma_s^2 + \tilde{\sigma}_s^2}\right]\right). \tag{22}$$

Plugging in all of the above expressions and simplifying yields

$$P(g = 1|x_T, x_D) \approx \frac{1}{1 + \frac{P(g=2)}{P(g=1)} \exp\left(\frac{1}{4}\left[\frac{(\hat{v}_T - \hat{v}_D)^2}{\sigma_v^2 + \tilde{\sigma}_v^2} + \frac{(\hat{s}_T - \hat{s}_D)^2}{\sigma_s^2 + \tilde{\sigma}_s^2}\right]\right)}. \tag{23}$$

Observe three properties of this result. First, the probability that the target and the decoy are grouped together is decreasing in the distance between them, as we expect from a measure of similarity. Its functional form traces out a logistic gradient. Second, the gradient is shallower when group-level dispersion and observation noise are high. Intuitively, this happens because variation makes it more plausible that items which appear quite different are actually generated by the same underlying process. The distance between items is measured in terms of how many standard deviations apart they are, providing a domain-general metric. Third, the assessment is tied to the prior probability that items come from the same group. This prior probability expresses the agent's default belief that two items were generated by the same process when nothing is known about their attributes.

**Model Extensions**

Although we made a number of assumptions for clarity, many of them can be relaxed to encompass a wider variety of environments. Many extensions are relatively standard in the Bayesian framework, such as allowing options to have different signal precisions or numbers of signals (Berger, 1985; Gelman et al., 2013). Here are several further ways to extend the model and broaden its applicability.

1. Different objective functions can be used. In particular, value could come from the product of the attributes as with expected utility or Cobb-Douglas preferences, $v_i = x_{i,1}^{\omega_1} x_{i,2}^{\omega_2}$ or equivalently $log(v_i) = \omega_1 log(x_{i,1}) + \omega_2 log(x_{i,2})$. This functional form with $\omega_1 = \omega_2 = 1$ could also capture perceptual judgments of rectangle size based on the product of width and height. Roughly following Shenoy and Yu (2013), the objective function in the Cobb-Douglas case can be inverted to find attribute magnitudes such that $log(\hat{v}_i) = \omega_1 log(x_{i,1}) + \omega_2 log(x_{i,2})$ and $\hat{s}_i log(\hat{v}_i) = \omega_1 log(x_{i,1}) - \omega_2 log(x_{i,2})$, which yields $x_{i,1} = \hat{v}_i^{\frac{1+\hat{s}_i}{2\omega_1}}$ and $x_{i,2} = \hat{v}_i^{\frac{1-\hat{s}_i}{2\omega_2}}$. Note that $\tilde{v}_i$, $v_i$ and $\hat{v}_i$ here should be positive. They might be modeled by gamma distributions as in Shenoy and Yu (2013), or by normal distributions truncated below at zero. When these distributions have a sufficiently large mean (relative to variance), they are approximately Gaussian; hence, when values are not too close to the zero bound, this setting should be approximated by the Gaussian structure assumed in the unbounded additive case, leading to similar conclusions.

2. Any number of options may be considered. The calculations of the posterior mean and group membership can be altered accordingly. Redefine $g = (g_1, \ldots, g_N)$ to represent the grouping of all $N$ options, where $g_i$ is an index of option $i$'s group membership taking on integer value $k$ between 1 and $K$. Then the posterior mean value of option $i$'s group becomes the average of signal values of all options in that group, $E[\tilde{v}_i | x_{j:g_j=g_i}] = \frac{1}{|\{j:g_j=g_i\}|} \sum_{j:g_j=g_i} \hat{v}_j$. The computation of group membership

must sum over all combinations of options, so $P(g \mid x_j \forall j) = \frac{P(x_j \forall j|g)P(g)}{\sum_{g'} P(x_j \forall j|g')P(g')}$ where $g'$ iterates over all possible groupings. We may also suppose that the prior probability of a group assignment is given by the Chinese restaurant process (Aldous, 1985; Gershman & Blei, 2012), a nonparametric Bayesian model which has the capacity to accommodate any number of groups. The prior probability of grouping $g$ is then given by $P(g \mid \alpha, N) = \frac{\alpha^K \Gamma(\alpha) \prod_{k=1}^K \Gamma(|\{j:g_j=k\}|)}{\Gamma(N+\alpha)}$ where $\Gamma(\cdot)$ is the gamma function and $\alpha$ is the concentration parameter that represents how strongly items tend to cluster into large groups. When $\alpha = 0$, all items will be part of the same group, while in the limit as $\alpha \to \infty$, each item will constitute its own unique group. In the two-item case above, $P(g = 2) = \frac{\alpha}{1+\alpha}$ and $P(g = 1) = \frac{1}{1+\alpha}$, so $\frac{P(g=2)}{P(g=1)} = \alpha$.

3. Any number of latent properties or observable attributes may be considered. Let $\mathbf{x}_i$ be an $M$-dimensional vector of observable attributes with $m$th element $x_{i,m}$, let $\mathbf{y}_i$ be an $L$-dimensional vector of latent properties (one of which represents value) with $\ell$th element $y_{i,\ell}$, and let $\hat{\mathbf{y}}_i = \mathbf{y}_i + \boldsymbol{\varepsilon}_i$ be a noise-corrupted signal of the latent properties where $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I})$ denotes noise with an $L$-dimensional vector of variances $\boldsymbol{\sigma}^2$ that has $\ell$th element $\sigma_\ell^2$. Similarly, suppose $\mathbf{y}_i \sim \mathcal{N}(\tilde{\mathbf{y}}_i, \tilde{\boldsymbol{\sigma}}^2\mathbf{I})$ with $L$-dimensional mean and variance vectors. Assume that $\mathbf{x}_i = \mathbf{Z}\hat{\mathbf{y}}_i$ for some $M \times L$ matrix $\mathbf{Z}$ that transforms latent properties into observable attributes, and hence $\hat{\mathbf{y}}_i = \mathbf{Z}^{-1}\mathbf{x}_i$. For example, in the case laid out earlier with $M = L = 2$, $\mathbf{y}_i = \left[\begin{smallmatrix} v_i \\ s_i \end{smallmatrix}\right]$, $\hat{\mathbf{y}}_i = \left[\begin{smallmatrix} \hat{v}_i \\ \hat{s}_i \end{smallmatrix}\right]$, $\mathbf{Z} = \left[\begin{smallmatrix} \frac{1}{2\omega_1} & \frac{1}{2\omega_2} \\ \frac{1}{2\omega_1} & -\frac{1}{2\omega_2} \end{smallmatrix}\right]$, and $\mathbf{Z}^{-1} = \left[\begin{smallmatrix} \omega_1 & \omega_2 \\ \omega_1 & -\omega_2 \end{smallmatrix}\right]$. Then $E[y_{i,\ell}|\mathbf{x}_j \forall j] = \hat{y}_{i,\ell} - \frac{\sigma_\ell^2}{\sigma_\ell^2 + \tilde{\sigma}_\ell^2}(\hat{y}_{i,\ell} - \bar{y}_\ell)$ where $\bar{y}_\ell$ is the sample average value of $\hat{y}_{j,\ell}$ across members in the same group as $i$.

4. An informative hyperprior over values can be incorporated. People may come into a task with an initial bias. For example, Viscusi (1989) proposed that an individual's predisposition for optimism or pessimism could color their perceptions of risky gambles. This bias at a higher level might be a fruitful way to model individual differences. Suppose that the decision maker holds a hyperprior over values such that

$\tilde{v} \sim \mathcal{N}(\tilde{\tilde{v}}, \tilde{\tilde{\sigma}}_v^2)$. Then the Bayesian estimates of the group mean become

$$E[\tilde{v}_T \mid x_T] = \left(\frac{\tilde{\tilde{\sigma}}_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2 + \tilde{\tilde{\sigma}}_v^2}\right) \hat{v}_T + \left(\frac{\sigma_v^2 + \tilde{\sigma}_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2 + \tilde{\tilde{\sigma}}_v^2}\right) \tilde{\tilde{v}} \tag{24}$$

$$E[\tilde{v}_T \mid x_T, x_D] = \left(\frac{\tilde{\tilde{\sigma}}_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2 + 2\tilde{\tilde{\sigma}}_v^2}\right) (\hat{v}_T + \hat{v}_D) + \left(\frac{\sigma_v^2 + \tilde{\sigma}_v^2}{\sigma_v^2 + \tilde{\sigma}_v^2 + 2\tilde{\tilde{\sigma}}_v^2}\right) \tilde{\tilde{v}}. \tag{25}$$

These estimates and the resulting estimates of option value are biased toward the hyperprior mean. Note that the original solution with an uninformative hyperprior is recovered as $\tilde{\tilde{\sigma}}_v^2 \to \infty$.

5. Variation may be treated as uncertain. Although we supposed the degree of variation along each dimension at each level was known by the decision maker, people might instead learn about variances based on how items cluster together (Anderson, 1991; André, Reinholtz, & De Langhe, 2021; Navarro & Kemp, 2017). This learning would affect the strength of assimilation and inference about latent groups. Such local calibration could help explain how context effects may manifest differently in disparate environments and occur at all scales.

6. Different attribute types can be analyzed. This flexibility lets us define similarity and apply the tainting hypothesis in a wide range of environments. For instance, options may consist of features which are either categorically present or absent, so options can be represented as binary vectors. In this setting, Kemp et al. (2005) have shown how Tversky's (1977) contrast model can be recovered if the feature vectors are generated by a beta-Bernoulli model, demonstrating a link between latent group inference and classic theories of similarity.

7. Different beliefs about latent group structure may be explored. Such versatility can be useful when describing how judgments may differ across economic markets or perceptual modalities. For example, people might think that in markets, companies strive to create diversified product lines. This belief could imply that products which have overly similar attributes are actually likely to have come from different brands.

In theory, this might even produce an attraction-like effect, as the decoy could be grouped together with the competitor to the benefit of the target. Beliefs about market structure could also emerge from strategic equilibria that might affect the information content of the choice set (e.g., Kamenica, 2008).

## Applications

In this section, we discuss several factors that have been observed to modulate context dependence, and show how our proposed mechanism might yield this systematic variation.

### Decisions from Experience

A body of work studies the difference between decisions when the attributes of options are explicitly described versus when they must be learned from experience (e.g., Hertwig, Barron, Weber, & Erev, 2004; Hertwig & Erev, 2009). For example, an agent may be told that a gamble offers a $10 payoff with a 50% probability, or they may acquire a sample of realized outcomes from which they must estimate these quantities. The resulting estimates may be noisy and distorted in various ways. Ert and Lejarraga (2018) conducted repeated-choice tasks in which participants were either provided with explicit descriptions of risky gambles or had to learn their properties from experience, and observed a repulsion effect only in the latter case. Spektor et al. (2019) also observed a repulsion effect in a learning-from-experience task. To accommodate this phenomenon, they extended a basic reinforcement learning model such that representations of options could interact with each other in the updating process. Their augmentation consisted of a mechanism in which similar options (like the target and decoy) tend to inhibit each other, benefiting other options (like the competitor).

Decisions from experience rely on noisier estimates of value (e.g., Fox & Hadar, 2006). Through the lens of our theory, this noise should be counteracted by Bayesian regularization toward other options from the same group. Hence, the evaluation of the

target should be more subject to repulsion in decisions from experience. This mechanism is visualized in Figure 5.

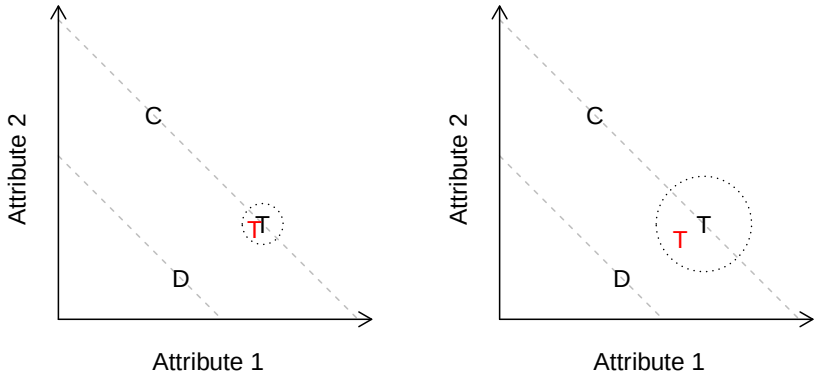


*Figure 5*. Illustration of the repulsion effect in decisions from experience (right) versus decisions from description (left).

In a dynamic setting, our proposed mechanism is related to hierarchical Bayesian models of reinforcement learning that invoke latent statistical structure. For example, when consumers learn about different products from the same brand or category, there can be spillover because experience with one product is informative about the quality of the others. Evidence for such correlated learning of option values has been found in abstract reward learning tasks (Acuña & Schrater, 2010; Schulz, Franklin, & Gershman, 2020; Wu, Schulz, Speekenbrink, Nelson, & Meder, 2018) and the exploration patterns of fishermen (Marcoul & Weninger, 2008), as well as the purchase behavior of consumers (Ching et al., 2013; Erdem, 1998; Erdem & Chang, 2012; Sridhar et al., 2012). An inferior decoy would thus reduce the appraised value of the associated target. Note that many previous models encode higher-order relationships in terms of fixed correlation parameters connecting different options. Our model allows these correlations to be derived based on the agent's probabilistic assessment of whether the items were generated by the same latent process.

### The Double Decoy Effect

Typical studies on the decoy effect include only a single decoy. However, Daviet and Webb (2020) conducted experiments with preferential choices in which a second decoy superior to the first one was added to the choice set. They observed that the addition of the second decoy made the target less appealing, cancelling out the attraction effect. As the second decoy does not extend the range of attributes beyond the first, this finding is difficult to account for with traditional models of range normalization (e.g., Soltani, De Martino, & Camerer, 2012; Volkmann, 1951). Daviet and Webb (2020) explained the data with a model in which attribute values are compared and normalized in a pairwise fashion, and they used hierarchical Bayesian statistical techniques to accommodate individual heterogeneity in model parameters.

We provide an alternative (though not mutually exclusive) explanation. In light of our theory, the second decoy provides further information about the group value, and furnishes an additional signal that the target may be worse than it appears. This extra signal should thus further bias the target's estimate downward. It might also increase the tendency to group the target and decoys together because its presence in between the target and the first decoy may contribute evidence that they are part of the same cluster. See Figure 6 for illustration.

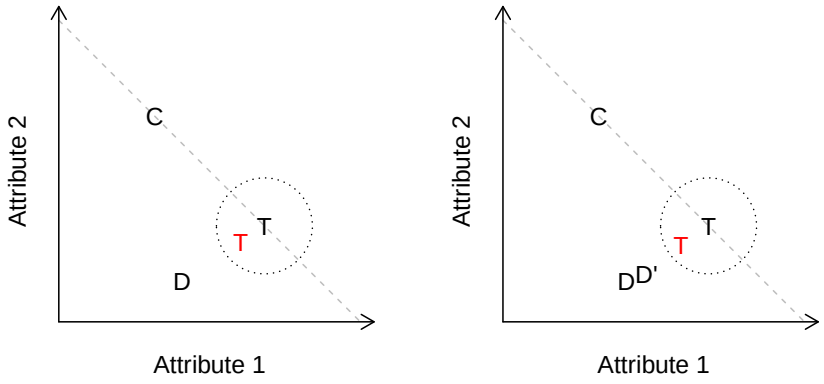


*Figure 6*. Illustration of the double decoy effect. D' represents the second decoy.

Similar mechanisms have been found to describe a variety of judgments (Austerweil et al., 2015). For example, Bayesian models of latent structure can capture the flexibility and speed with which arbitrary social groups and subgroups can be constructed in ways that simple dyadic similarity models cannot, as well as the resulting patterns of interpersonal judgment (Gershman & Cikara, 2020). The introduction of an extra person with attributes in between two others has been found to increase the tendency to group those two together and the corresponding degree of social influence (Gershman et al., 2017; Lau et al., 2018).

### The Phantom Decoy Effect

Phantom decoys, different from standard decoys, are typically superior to the target option but unavailable at the time of choice. Despite being dominant rather than dominated, they have been found to increase preference for the target in economic tasks (e.g., Highhouse, 1996; Pettibone & Wedell, 2000, 2007; Pratkanis & Farquhar, 1992; Scarpi, 2011; Scarpi & Pizzi, 2013). Farquhar and Pratkanis (1993) recognized that phantom alternatives may contain information that helps to contextualize the value of other options and can thereby facilitate choice (Fischhoff, Slovic, & Lichtenstein, 1980, p. 124). Our theory provides a formalization of this idea.

Notice that, aside from the unavailability of the decoy, this phantom decoy effect mirrors the repulsion effect. In the former case, a superior decoy favors the option it dominates; in the latter case, an inferior decoy undermines the option that dominates it. Figure 7 illustrates the relationship. In our Experiment 1b, described later, we demonstrate that the repulsion effect occurs even when the inferior decoy is a phantom.

One prominent account of the phantom decoy effect is the similarity-substitution hypothesis, according to which people default to the target option as an effort-saving strategy when the preferred decoy is not available because it is the most similar alternative (Pettibone & Wedell, 2000, 2007; Tversky, 1972). Our Bayesian account offers further nuance to this hypothesis. First, it provides an alternative adaptive rationale for

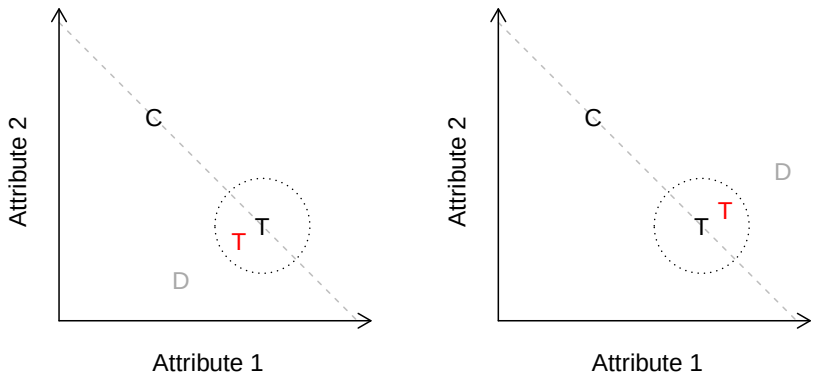


*Figure 7*. Illustration of the phantom decoy effect. The classic phantom decoy effect (right) is the mirror of the repulsion effect (left).

similarity-based judgment. Rather than effort minimization, the effect could stem from principles of inference under uncertainty. Second, it provides a refinement of what form similarity takes. Rather than a simple exponential function, similarity can manifest in different ways depending on the generative process assumed by the decision maker. Third, it integrates this similarity judgment into the overall effect of the decoy. This yields new implications for how the strength of the effect depends non-monotonically on the attribute distance between the decoy and the target, in contrast to previous hypotheses which predict monotonic relationships (Scarpi & Pizzi, 2013). We elaborate on this non-monotonicity in the next subsection.

### Non-Monotonic Effect of Decoy Distance

Decoys can be close to or far from the target in attribute space, and the strength of the repulsion effect appears to be modulated by this attribute distance. In consumer choice, Liao et al. (2020) found that distance has a U-shaped effect on the strength of the repulsion effect. That is, preference for the target first declines and then grows again as the decoy becomes more distant.[7]

[7] We do note that there have been mixed results in the literature. For example, Liao et al. (2020) found an inverse-U shaped effect of decoy distance in perceptual judgment. We speculate that this pattern could

This non-monotonic effect can occur in our model due to the conjunction of two factors determining the repulsion effect. First, when the decoy is moved farther away from the target, it constitutes a more negative signal. This effect grows linearly in the distance between the target and decoy. Second, the farther away the decoy is, the less likely it was generated by the same process as the target. This effect is decreasing at an approximately logistic rate. The product of these forces yields an initial enhancement of the repulsion effect (when the first force is stronger) followed by a decline (when the second force is stronger). The result is illustrated in Figure 8.

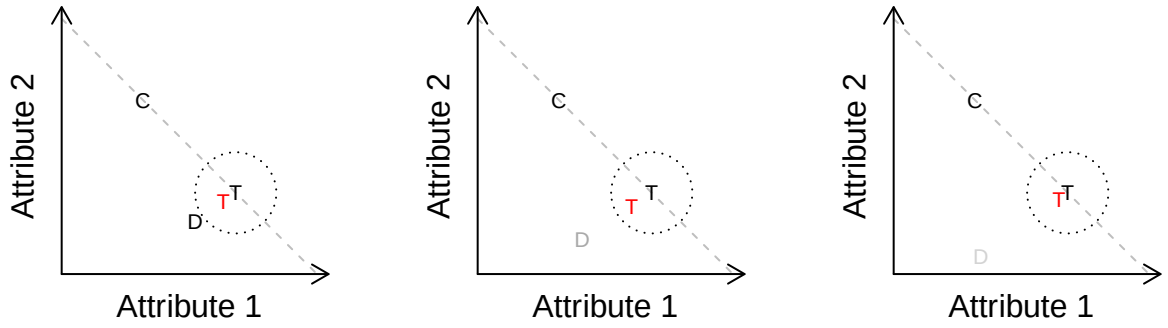


*Figure 8*. Illustration of the non-monotonic effect of decoy distance on the repulsion effect. Grayness represents the probability that the decoy is judged to be from the same group as the target.

Such a mechanism has been found to capture analogous patterns in multisensory perception (Cao et al., 2019; Körding et al., 2007; Sato et al., 2007; Shams & Beierholm,

arise from an analogous similarity-based mechanism applied to the attraction effect. If there were some baseline attraction force (call it $A$) that depended on how inferior the decoy was and that scaled by the probability that the target and decoy were from the same group, then an inverse-U effect of decoy distance would occur when $(A - \frac{\sigma_v^2}{\sigma_v^2+\tilde{\sigma}_v^2})(\frac{\hat{v}_T-\hat{v}_D}{2})P(g=1|x)$ is positive, or when $A > \frac{\sigma_v^2}{\sigma_v^2+\tilde{\sigma}_v^2}$. Indeed, several have argued that the attraction effect may be modulated by perceived similarity between the target and decoy (Izakson, Zeevi, & Levy, 2020; Mishra, Umesh, & Stem, 1993). These perspectives might be sharpened by the modeling of similarity in terms of latent group inference as we do here. We leave this as an open possibility beyond the scope of the present work, as we remain agnostic about the causes of the attraction effect in this paper.

2010). People draw causal inferences about whether two sensory cues originate from the same source (i.e., from the same location) or different sources, based on how concordant the cues are and the level of uncertainty in each one. This is why successful ventriloquists, for example, maximize the synchrony between the puppet's facial movements and the sound of their speech. When cues are close together, they are likely caused by the same source, and the best estimate of the source's location falls in between the cues (closer to the more precise signal). When the cues are far apart, they are likely caused by different sources, and the best estimates of the source locations are close to each cue separately. The estimated location of a cue's source is thus nonlinear in the distance between the cues. A similar observation has been made regarding anchoring and hindsight bias (Wilson et al., 2021). Knowledge of an outcome serves as an anchor that pulls recollections toward it. This bias depends non-monotonically on the distance between the anchor and the original estimate (Hardt & Pohl, 2003), which can be explained by the combination of the information provided by the anchor and the reduction in the anchor's plausibility as it increasingly deviates from expectations.

## Experiments

Our theory makes novel predictions about the ingredients needed for the repulsion effect to occur. We conducted new experiments to test some of these predictions. First, the decision maker must believe the target and the decoy come from the same group. Second, because the decoy is providing information about the target, its presence may still have an effect even when it cannot be chosen. Third, observable attributes must be noisy signals of underlying value, and the values of the target and decoy must be meaningfully correlated. Accordingly, we ran consumer choice experiments in which we varied group membership, both when the decoy was available (Experiment 1a) and when it was an unavailable phantom (Experiment 1b), and we ran value estimation tasks in which we varied statistical properties of the environment (Experiment 2a) and made the group structure implicit

(Experiment 2b). To foreshadow our results, we found that repulsion effects indeed emerged more strongly when the target and decoy were grouped together (even when the decoy was unavailable), when attribute noise was high and value dispersion was low, and when group membership could be more clearly inferred from stimulus attributes.

## Experiment 1a

In Experiment 1a, we explicitly manipulated the group structure underlying the options in a consumer choice setting, either saying that the target and decoy came from the same group, or that the three options came from different groups. This allowed us to test our primary hypothesis that the repulsion effect would occur when people believed that the target and decoy were generated by a similar process. If the repulsion effect did not depend on this belief, there should be no differences in participants' preferences across the two conditions.

**Participants.** 449 participants were recruited from Amazon Mechanical Turk. They received a $0.50 participation fee for completing the experiment. To ensure data quality, participants answered an attention check question at the end of the experiment, which can be seen in the Supplemental Materials. A total of 372 participants passed the attention check, and are included in the analyses below. The sample size was determined based on a power calculation which indicated that it would provide 90% power to detect a 3-point shift (corresponding to a Cohen's $d$ of about 0.1 according to pretests) at the 5% significance level assuming 20% exclusion from participants failing the attention check and 2% loss of data from technical glitches, when pooling across all stimuli. The experiment was approved by the MIT Committee on the Use of Humans as Experimental Subjects. All participants gave informed consent prior to their participation.

**Procedure.** Each participant made three hypothetical consumer choices based on stimuli borrowed from existing paradigms: Italian restaurants (adapted from Sen 1998), cans of orange juice (adapted from Ratneshwar, Shocker, and Stewart 1987), and lottery

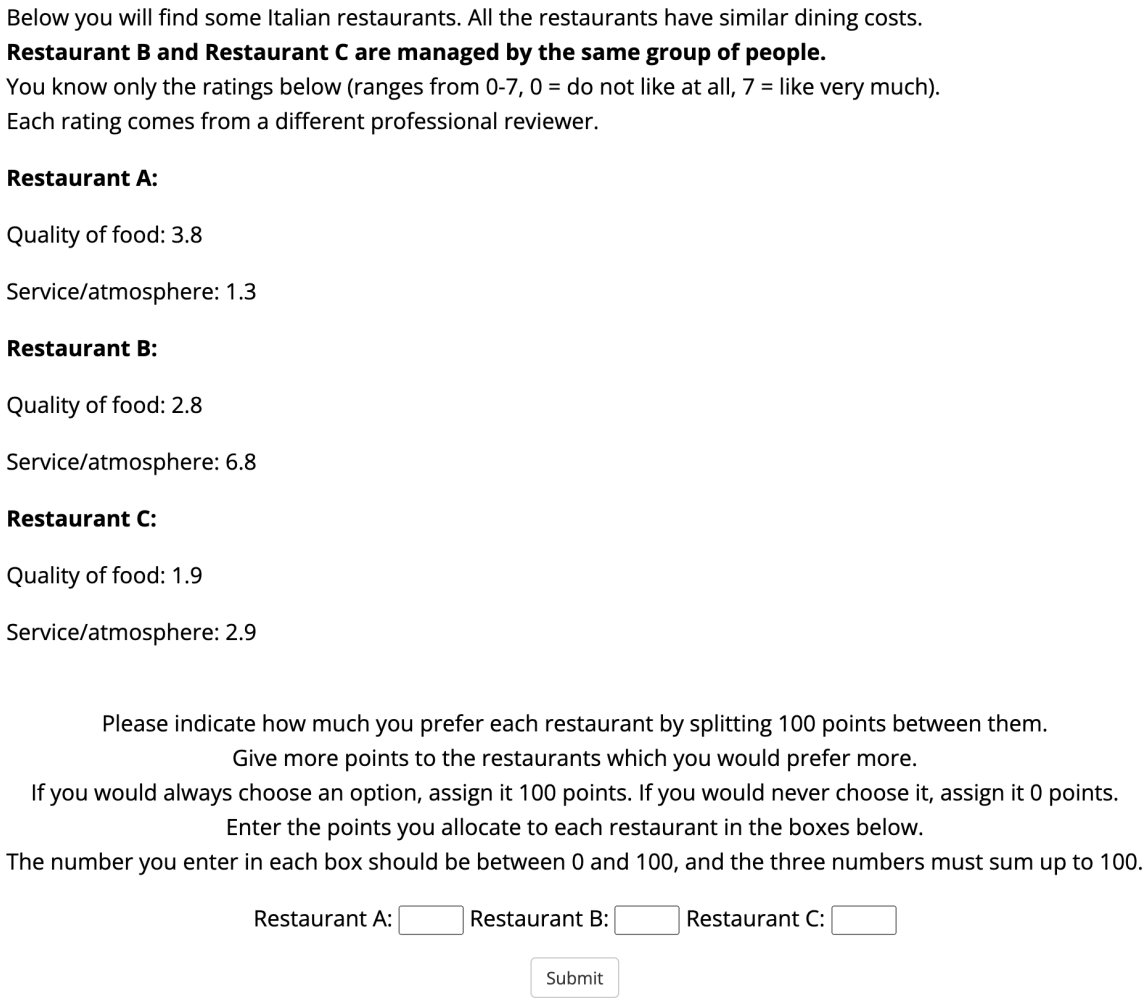


*Figure 9*. Screenshot of consumer choice task from Experiment 1a. Participants were told either that all three options were from different groups or that the target and decoy came from the same group.

tickets (adapted from Huber et al. 1982). Each of these three choices consisted of three different options, with attributes that corresponded to a target, a competitor, and a decoy. The stimulus attributes are listed in the Supplemental Materials. Participants indicated the strength of their preference for the options by splitting 100 points between them. Figure 9 shows a screenshot of the task.

We used a between-subjects design in which participants were randomly assigned to one of two conditions. Participants in the "same group" condition were explicitly told in all

three choices that the target and decoy came from the same brand or were managed by the same group of people; participants in the "different groups" condition were told that all three options came from different brands or were managed by different groups of people. This text was emphasized to ensure participants were aware of it. They had unlimited time to answer the questions.

To counterbalance the target and competitor identity, for each question we designed a separate decoy for each side (randomly assigned to participants), inferior in one dimension. This allows us to eliminate the effect of target identity by pooling together data across both decoy sides. Each option had two attributes (e.g., Quality and Price). There were two versions of each stimulus that were roughly matched, a qualitative version with verbally described attributes and a quantitative version with numerically described attributes. These versions were randomized within each participant, meaning every participant would face one choice for each of the three stimulus types but could have a mix of quantitative and qualitative versions of each. Pretests were conducted to avoid floor and ceiling effects. The order of choice problems was counterbalanced. The results presented below pool across all three questions and both attribute versions.

**Exclusions.** We excluded 77 participants for failing the end-of-task attention check.

**Results.** When all three options came from different groups, there was no significant difference between the points assigned to the target and the competitor [45.4 vs 45.4; $t(497) = -0.03$, $p = .974$]; the decoy seemed to have no effect. However, a repulsion effect emerged when the target and decoy came from the same group, as points assigned to the target were significantly lower than points assigned to the competitor [40.6 vs 48.0; $t(617) = -3.45$, $p < .001$]. This difference corresponded to a significant drop in points assigned to the target across conditions [$t(1040.3) = -2.92$, $p = .004$], as predicted by the theory. As another hallmark of the theory, the difference was accompanied by an increase in the points assigned to the decoy [9.2 vs 11.4; $t(1098) = 2.51$, $p = .012$]. Figure 10 visualizes the pattern.

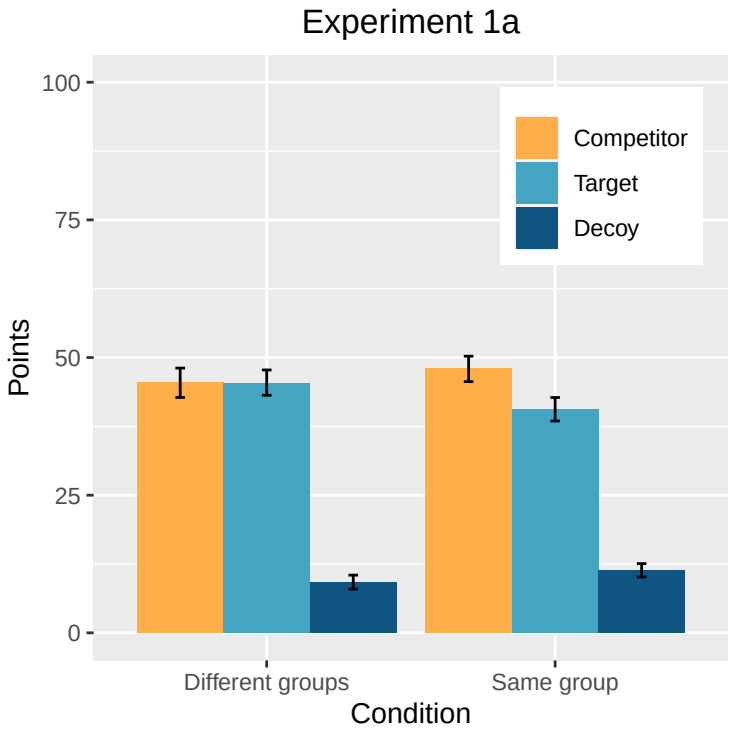


*Figure 10*. Results from Experiment 1a. Point allocations indicating strength of preference are shown with 95% confidence intervals.

## Experiment 1b

The aim of Experiment 1b was to determine whether the results of Experiment 1a could be extended to a phantom decoy paradigm, in which the inferior decoy was present but not available to be chosen.

**Participants.** 449 participants were recruited from Amazon Mechanical Turk. They received a $0.50 participation fee for completing the experiment. To ensure data quality, participants answered an attention check question at the end of the experiment. A total of 410 participants passed the attention check, and are included in the analyses below. The sample size was determined similarly to Experiment 1a. The experiment was approved by the MIT Committee on the Use of Humans as Experimental Subjects. All participants gave informed consent prior to their participation.

**Procedure.** The procedure was similar to that of Experiment 1a, except that participants allocated points only to the target and competitor. Thus, the group structure was manipulated while the decoy did not directly enter into the preference comparison.

**Exclusions.** We excluded 39 participants for failing the end-of-task attention check.

**Results.** Similar to results of Experiment 1a, when the target, decoy, and competitor came from different groups, the number of points assigned to the target was not significantly different from points assigned to the competitor [50.6 vs 49.4; $t(632) = 0.47$, $p = .640$]. However, fewer than half the points were assigned to the target in the condition where the target and decoy came from the same group, indicating a significant repulsion effect [46.4 vs 53.6; $t(596) = -2.89$, $p = .004$]. This difference corresponded to a significant drop in points assigned to the target across conditions [$t(1224.7) = -2.39$, $p = .017$]. Figure 11 visualizes the pattern.

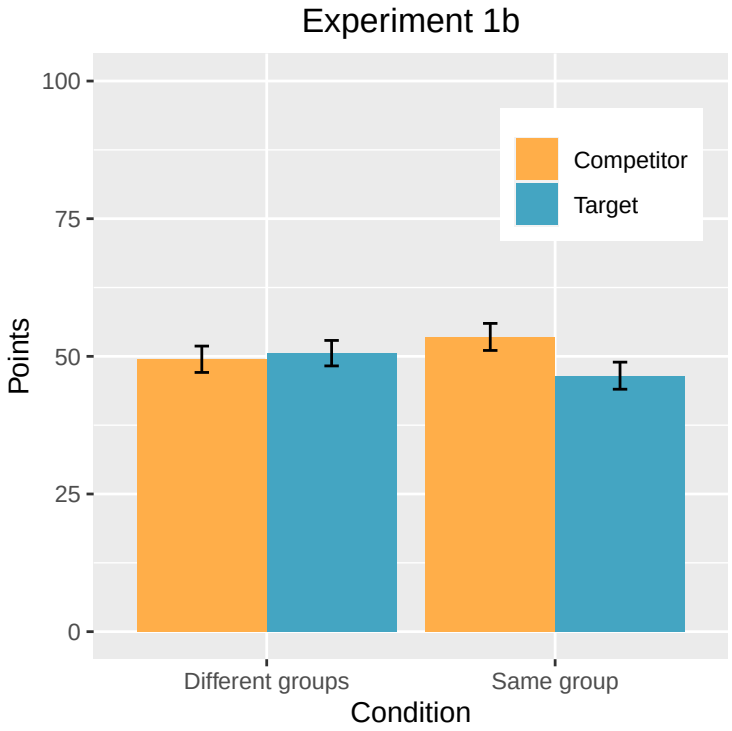


*Figure 11*. Results from Experiment 1b. Point allocations indicating strength of preference are shown with 95% confidence intervals.

## Discussion

Participants preferred the target option less when it came from the same group as the inferior decoy, leading to a repulsion effect. Preference for the decoy grew in tandem. The decline in preference for the target occurred even when preference for the decoy was not elicited.

## Experiment 2a

The goal of Experiment 2a was to test whether the repulsion effect varies depending on uncertainty in the environment. Specifically, the effect should be stronger when attributes are noisier signals of value and there is little dispersion in the values of members from the same group, meaning there is a high noise-to-dispersion ratio. To precisely control the statistical properties of the environment and measure how this influences the formation of beliefs, we turned to a different paradigm based on value estimation. This task was also incentivized and repeatable, so the same participant could be posed many trials.

**Participants.** 80 participants were recruited from Amazon Mechanical Turk. They received a base pay of $1.00 for participation and a performance bonus up to $3.50. To ensure that participants understood the rules, they completed a comprehension check before they were allowed to move on to the main task. Those who failed the check were given the option to try again or review the instructions again, until they correctly answered all the comprehension check questions. The experiment was approved by the MIT Committee on the Use of Humans as Experimental Subjects. All participants gave informed consent prior to their participation.

**Procedure.** Participants were instructed to imagine themselves as gold miners in the Wild West, borrowing task aesthetics and stimuli from Dorfman, Bhui, Hughes, and Gershman (2019). On each trial, they were shown three mines, and had to provide their best guess of how much gold would be found in each one. Noisy signals of these true values were presented as test samples to inform the estimates, as seen in Figure 12. Participants

were paid based on the absolute error between their guesses and the true values. Two of the mines were shown to be from the same mountain, leading to correlated values, while the third was from a different mountain. We refer to a given one of the first two mines as the "target mine," the other mine from the same mountain as the "decoy mine" (though it is not necessarily inferior to the target), and the mine from a different mountain as the "competitor mine." The equivalent of the repulsion effect in this task occurs when the estimated value of the target mine is biased by the signal of the decoy mine.

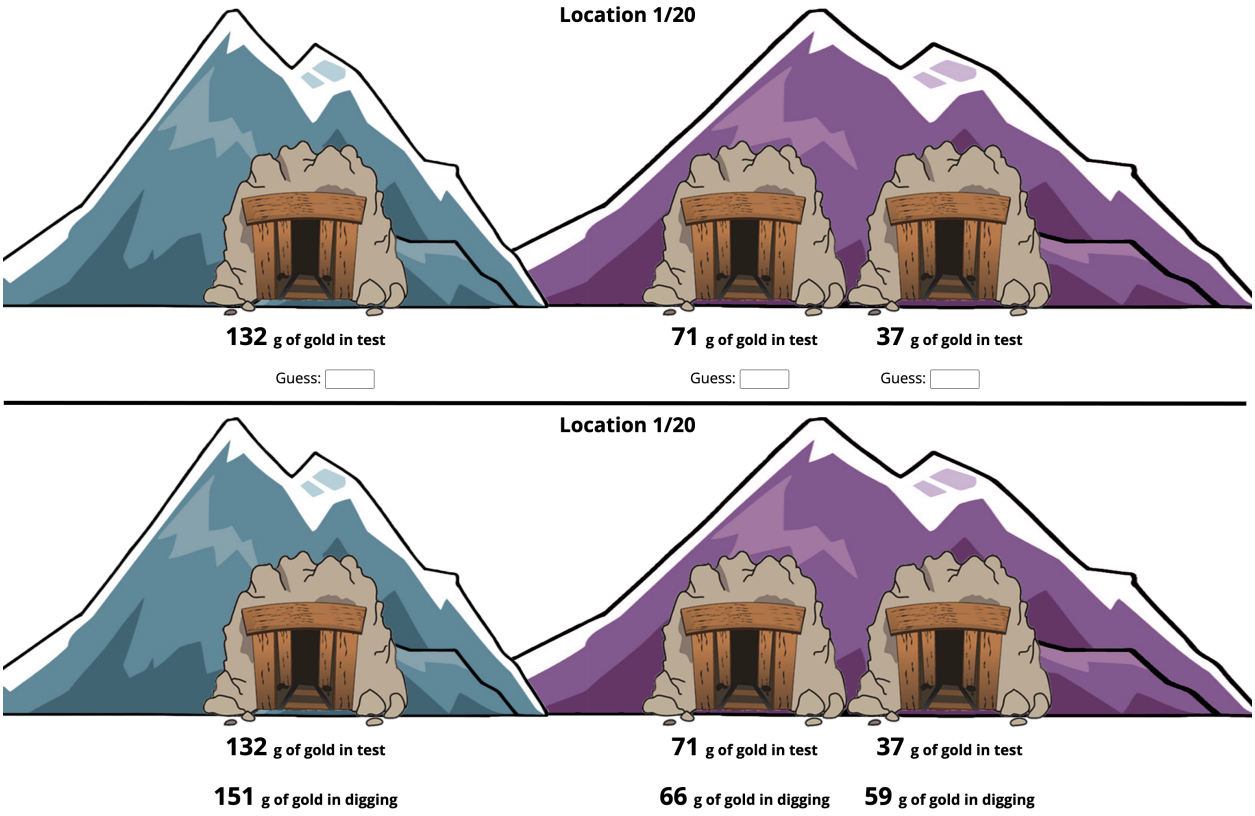


*Figure 12*. Screenshot of value estimation task from Experiment 2a. The mine on the left is from one mountain (competitor mine) and the two mines on the right are from another mountain (target mine and decoy mine). Participants were presented with the test samples and reported their guesses of the true values of each mine. They then received feedback on the true values.

All of the values were drawn from a hierarchical distribution similar to the generative

model in our theory, with independent draws across trials. The expected values of each mountain were drawn from a discrete uniform distribution ranging between 50 and 150 in increments of 1. The value of each mine was drawn from a Gaussian distribution with mean equal to the value of the mountain it belonged to and variance (i.e., dispersion) that depended on the experimental condition. Thus, these values were positively correlated for the two mines from the same mountain. The noisy signals were drawn from Gaussian distributions with mean equal to the mine value and variance (i.e., noise) that again depended on the experimental condition. Values were redrawn in the rare case they fell below 0.

Participants completed two blocks of 20 trials each. In one block, the standard deviation of the mine values was 15 while the standard deviation of the noisy signal was 5, yielding a low noise-to-dispersion ratio. In the other block, the two standard deviations were reversed, yielding a high noise-to-dispersion ratio. The order of blocks was counterbalanced. Participants received feedback on the true mine values after each trial. To help them become acquainted with the value distributions, we also showed them 10 training examples before each block.

If the repulsion effect is modulated by the noise-to-dispersion ratio as the theory predicts, the experimental manipulation should affect how participants' guesses rely on the test sample from each mine. When the ratio is higher, participants' estimates of the target mine should rely relatively less on the target mine's own test sample and more on the test sample of the decoy mine. Assuming they know the generative process well enough, they should not rely on the test sample of the competitor mine in any case.

**Exclusions.** We excluded 11 outlier trials where participants' guesses were 500 away from the test sample, comprising 0.34% of the total trials.

**Results.** To quantify how participants integrated the various pieces of information, we analyzed the data using a random-effects regression which predicted participants' estimates of the mine values based on the experimental condition, the test values of all

three mines, and the interactions between condition and each of these signals, with subject-level random effects for all the regressors. This regression included only the pooled responses for the two mines from the same mountain, as estimates for the other mine would not be expected to have the same pattern.

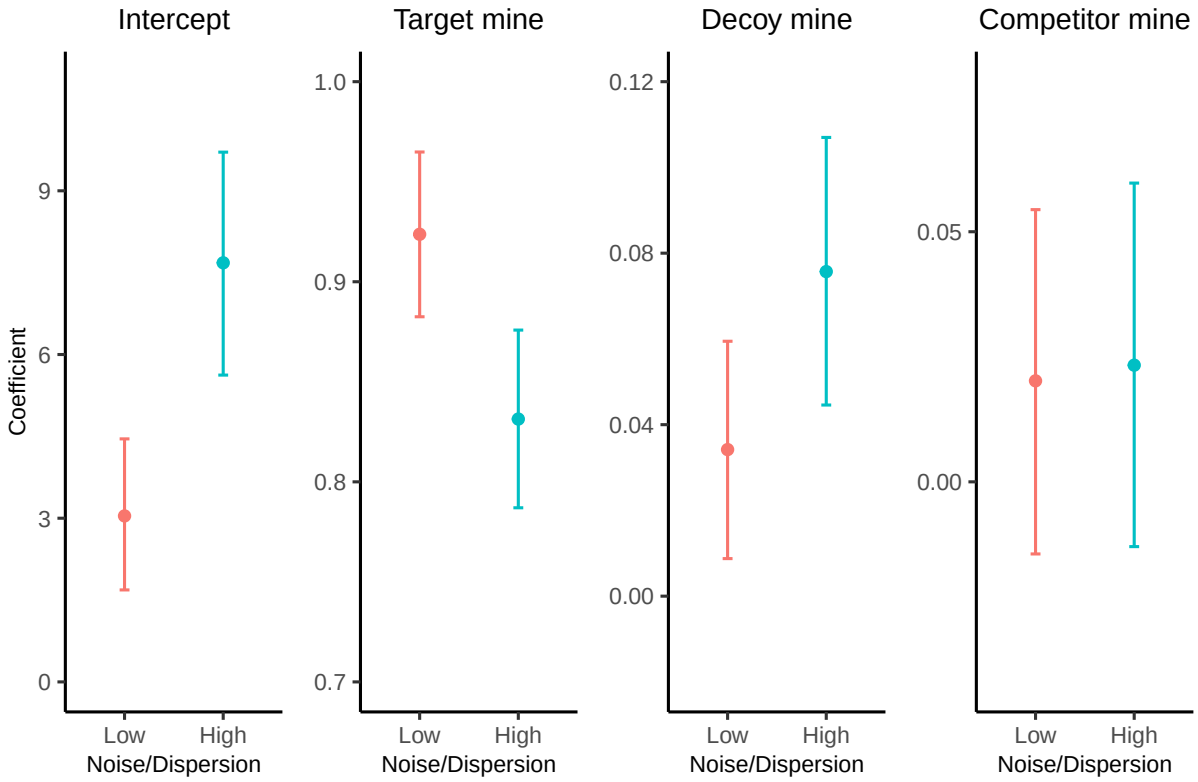


*Figure 13*. Results from Experiment 2a. Regression coefficients from subject-level random effects model are shown with 95% between-subjects confidence intervals. Noise/Dispersion = noise-to-dispersion ratio.

The regression results are visualized in Figure 13 and summarized in Table 1. When the noise-to-dispersion ratio was high, participants relied relatively less on the target mine's own signal and were influenced more by the decoy mine's signal, as predicted. This result was indicated by a statistically significant negative interaction between experimental condition (a dummy variable equal to 1 when the noise-to-dispersion ratio was high) and the target mine test value [$t(970.0) = -9.24$, $p < .001$] as well as a significant positive

Table 1

*Regression results from Experiment 2a.*

| | **Estimate** | **Standard error** | **Df** | **t value** | **p value** |
|---|---|---|---|---|---|
| (Intercept) | 3.043 | 0.819 | 122.3 | 3.717 | <.001 |
| Target mine | 0.924 | 0.021 | 84.5 | 45.060 | <.001 |
| Decoy mine | 0.034 | 0.013 | 99.8 | 2.574 | .012 |
| Competitor mine | 0.020 | 0.017 | 83.2 | 1.167 | .246 |
| Condition | 4.632 | 0.945 | 1667 | 4.903 | <.001 |
| Condition × Target mine | -0.092 | 0.010 | 970.0 | -9.237 | <.001 |
| Condition × Decoy mine | 0.042 | 0.010 | 2454 | 4.343 | <.001 |
| Condition × Competitor mine | 0.003 | 0.006 | 818.5 | 0.500 | .617 |

Note: Condition is a dummy variable which equals 1 in the high noise-to-dispersion ratio condition and 0 otherwise.

interaction between condition and the decoy mine test value [$t(2454) = 4.34$, $p < .001$]. No significant effect was observed for the interaction between condition and competitor mine test value [$t(818.5) = 0.50$, $p = .617$], and the competitor test value coefficients were not significantly different from 0 in either condition [$t(82.2) = 1.13$, $p = .261$ in the low noise-to-dispersion ratio condition, and $t(81.2) = 1.27$, $p = .207$ in the high noise-to-dispersion ratio condition], indicating no reliance on this unrelated signal. The intercept reflects reliance on the hyperprior, the distribution of mountain values, which had a mean of 100. The increase of the intercept in the high noise-to-dispersion ratio condition [$t(1667) = 4.90$, $p < .001$] implies that participants relied more on this hyperprior as well, also consistent with Bayesian principles.

## Experiment 2b

In previous experiments, we explicitly told participants the group structure. In Experiment 2b, we withheld information about group membership in order to examine how

latent group inference modulates the repulsion effect. To alter this inference, we showed participants the style of each item and varied the level of noise along this value-neutral dimension. This allowed us to keep the value distributions intact. We hypothesized that participants would infer group membership based on how similar the styles were in statistical terms, and would use that assessment to guide their guesses of mine values. When style noise is low, mines can be easily assigned to the correct mountains, and their estimated values should be more biased by the other mine from the same mountain.

**Participants.** 86 participants were recruited from Amazon Mechanical Turk. They received a base payment of $2.00 for participation and a performance bonus up to $7.00. To ensure that participants understood the rules, they completed a comprehension check before they were allowed to move on to the main task. Those who failed the comprehension check were given the option to try again or review the instructions again, until they correctly answered all the comprehension check questions. To make sure that they paid attention, participants also completed an attention check at the end of each block, which can be seen in the Supplemental Materials. Five participants failed at least one attention check. The experiment was approved by the MIT Committee on the Use of Humans as Experimental Subjects. All participants gave informed consent prior to their participation.

**Procedure.** Similar to the procedure of Experiment 2a, participants were in the role of miners in the Wild West, estimating the values of three mines. However, there were two key differences, as seen in Figure 14. First, participants were not told which mines belonged to which mountains. They only knew that two out of the three mines were from the same mountain and the remaining mine was from a different mountain. They had to infer which mine was from a different mountain and report their guess; this single response summarizes the entire group structure. Second, instead of mining for gold, participants mined for a mix of red ore and blue ore, which translated into total dollar values. In addition to test samples for each mine (i.e., noisy signals of the true values), they were shown the ratio of red to blue ore found in the sample. Although both kinds of ore had

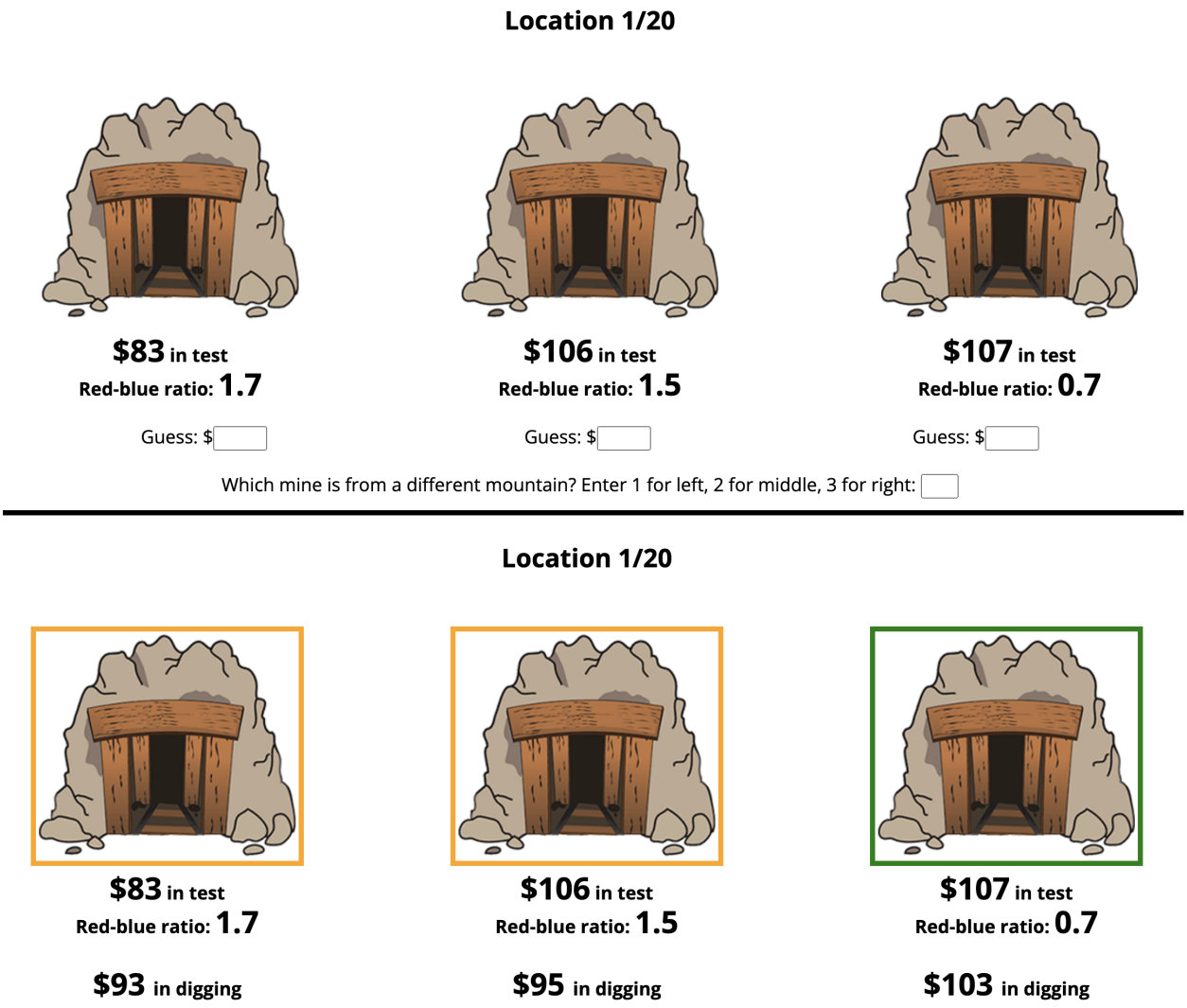


*Figure 14*. Screenshot of value estimation task from Experiment 2b. Participants were presented with the test samples and the red-to-blue ratios of each mine, without knowing which two mines belonged to the same mountain. They reported their guesses of the true values of each mine and which mine came from a different mountain. They then received feedback. Colored borders indicated mountain memberships; the mine with a green border was from one mountain (competitor mine), and the two mines with an orange border were from another mountain (target mine and decoy mine).

equal value, and participants were informed of this fact, this multi-attribute description could help assign mines to mountains. Participants again received bonus payments based

on the absolute error between their value estimates and the true values. Additionally, they received 2.5 cents for every time they correctly identified the mine which belonged to a different mountain.

The value of each mountain was drawn from a discrete uniform distribution ranging between 100 and 200 in increments of 1. The value of each mine was drawn from a Gaussian distribution with mean equal to the value of the mountain it belonged to and a standard deviation of 5. The noisy signals were drawn from Gaussian distributions with mean equal to the mine value and a standard deviation of 15. The styles of each mine were drawn from Gaussian distributions with a mean of $-20$ (mountain containing one mine) or $+20$ (mountain containing two mines) and variance depending on the experimental condition, translating into red-blue ratios averaging 0.8 and 1.3. All attribute magnitudes were redrawn in the rare case they fell below 0.

Participants completed two blocks of 20 trials each. In one block, the standard deviation of the mine styles was 2. In the other block, the standard deviation of the mine styles was 50. The order of the blocks was counterbalanced. Higher style noise would lead to more overlap between the ore ratio distributions of the two mountains. Participants received feedback on the true mine values and the membership of each mine after every trial. To help them become acquainted with the value distributions and the relationship between style and group membership, they were shown 15 training examples before each block. In these examples, they were first asked to make a guess of which mine was from a different mountain based on the test samples with noisy signals and styles, and then the true mine values and mountain assignments were revealed.

If participants relied on the styles of the mines to help infer latent group structure, then they should make more accurate inferences when style noise is low. Moreover, if the repulsion effect depends on the ability to correctly infer latent group structure, participants' estimates of target mine value should rely relatively less on the target's own test sample and more on the decoy mine's test sample when style noise is low.

**Exclusions.** One participant did not understand the task correctly and entered only 1, 2, or 3 in the guesses of total ore value, so we excluded this participant from our analysis. We also excluded the five participants who failed at least one attention check. Data was analyzed from the remaining 80 participants. We attempted to exclude trials where participants' guesses were 500 away from the test sample, as with Experiment 2a, but there were no such trials. We also removed 8 trials where the amount of blue ore was 0 leading to an undefined red-blue ratio.

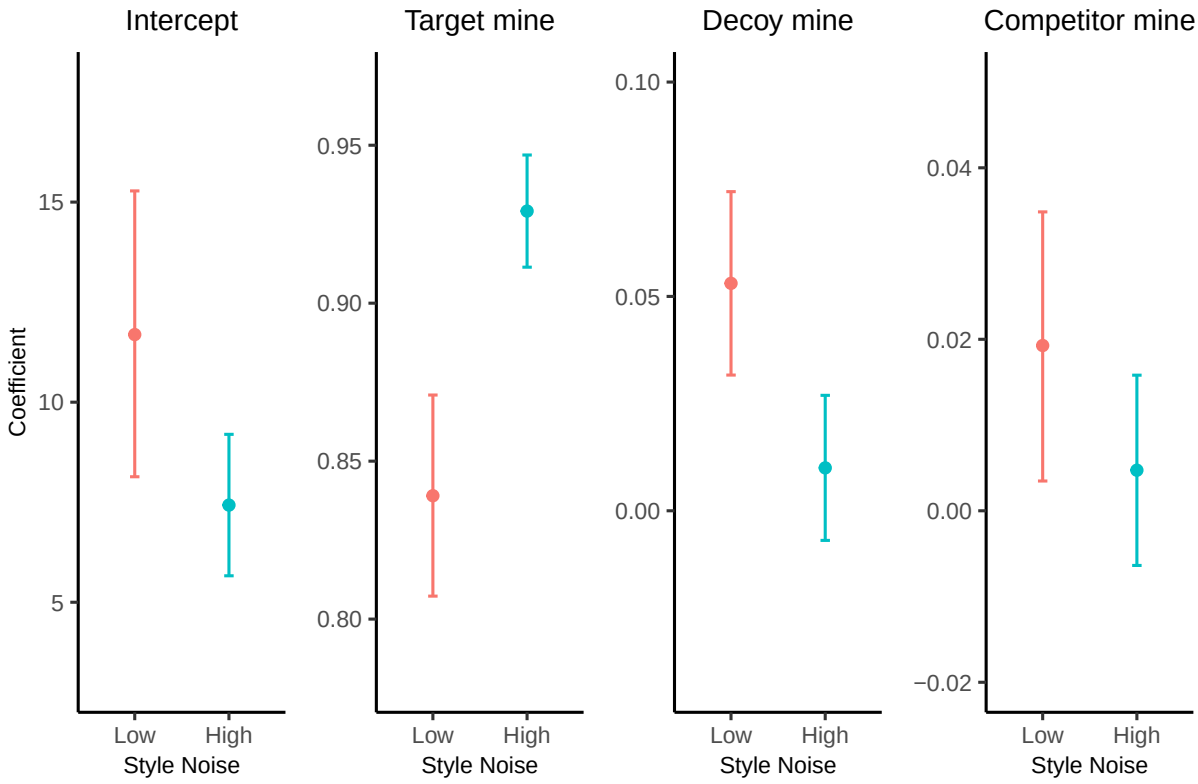


*Figure 15*. Results from Experiment 2b. Regression coefficients from subject-level random effects model are shown with 95% between-subjects confidence intervals.

**Results.** Participants more accurately identified the mines belonging to each mountain in the low style noise condition (92.5% correct) as compared to the high style noise condition (46.8% correct), a statistically significant difference according to a paired (within-subjects) t-test [$t(79) = -20.42$, $p < .001$].

Table 2

*Regression results from Experiment 2b.*

| | **Estimate** | **Standard error** | **Df** | **t value** | **p value** |
|---|---|---|---|---|---|
| (Intercept) | 11.564 | 3.720 | 389.4 | 3.109 | .002 |
| Target mine | 0.839 | 0.017 | 96.5 | 48.859 | <.001 |
| Decoy mine | 0.053 | 0.011 | 99.1 | 4.745 | <.001 |
| Competitor mine | 0.020 | 0.008 | 69.4 | 2.527 | .014 |
| Condition | -4.124 | 3.316 | 137.3 | -1.244 | .216 |
| Condition × Target mine | 0.091 | 0.016 | 151.4 | 5.680 | <.001 |
| Condition × Decoy mine | -0.044 | 0.014 | 218.9 | -3.185 | .002 |
| Condition × Competitor mine | -0.015 | 0.009 | 139.7 | -1.615 | .109 |

Note: Condition is a dummy variable which equals 1 in the high style noise condition and 0 otherwise.

As in Experiment 2a, we analyzed the data using a random-effects regression which predicted participants' guesses of the mine values based on the experimental condition (a dummy variable equaling 1 in the high style noise condition), the test values of all three mines, and the interactions between condition and each of the test values, along with subject-level random effects for all the regressors. This regression similarly included only the pooled responses for the two mines from the same mountain.

The regression results are shown in Figure 15 and Table 2. When the style noise was high, participants' estimates of the target mine value relied more on the target mine's own signal and were influenced less by the decoy mine's signal, as predicted. These results were indicated by a statistically significant positive interaction between condition and target mine test value [$t(151.4) = 5.68$, $p < .001$] and a significant negative interaction between condition and decoy mine test value [$t(218.9) = -3.19$, $p = .002$]. Though not predicted by the model in this setting, there was no significant interaction between condition and the test value of the competitor mine [$t(139.7) = -1.62$, $p = .109$] or the intercept

$[t(137.3) = -1.24,\ p = .216]$.

### Discussion

Participants' estimates of target mine value were biased toward the decoy mine signal, especially when noise was high and dispersion was low (meaning their attributes were worse predictors of true value and their true values were more correlated). Even when mines were not explicitly grouped together, target mine value estimates were biased more when low noise in the stylistic attribute allowed participants to more accurately infer group structure.

## Managerial Implications

The design of product lines and the construction of choice sets are central components of marketing strategy. While academic research and popular writing on these topics have traditionally emphasized the attraction effect (Ariely, 2008; Orhun, 2009), the existence of the repulsion effect means that a brand which introduces an inferior decoy option could damage itself. Our work clarifies the conditions under which this kind of backfire could happen.[8] Specifically, there must be uncertainty about the true value of the product (e.g., because the product or the consumer is new), uncertainty about the brand or category (e.g., because the brand or category is new), and the brand or category must be coherent (i.e., products must be thought to have similar properties). Further, because our model applies to perceptual judgment as well, the design attributes considered could include perceptual feature like product package size in addition to economic ones.

In this light, our work is evidently connected to research on umbrella branding. Strong empirical correlations have been found between consumer perceptions of products from the same brand (Bottomley & Holden, 2001; Erdem, 1998; Erdem & Chang, 2012; Erdem & Sun, 2002; Erdem & Swait, 1998; Erdem, Swait, & Valenzuela, 2006; Sullivan,

[8] Since we do not consider the attraction effect in this paper, our results might be considered necessary though not sufficient conditions for the repulsion effect to emerge on net.

1990). Theoretical models have shown how correlation between product values can be credibly signaled by umbrella branding (Keller, 2012; Miklós-Thal, 2012; Moorthy, 2012; Wernerfelt, 1988, 2012), helping to endogenize the underlying relationship we assume may be present. Our work also contributes to the broader literatures on consumer inference (Kardes, Posavac, Cronley, & Herr, 2008) and category-based inference (Loken, Barsalou, & Joiner, 2008). For example, our results formally recapitulate the findings of Levin and Levin (2000) that inferences about the qualities of an incompletely described target product are guided by the qualities of a second, linked product that is more well-defined.

The decoy's detrimental impact in the repulsion effect raises the question of whether it could occur in a competitive scenario. This could be possible under the conditions outlined above, and might have happened in the cola wars during the early 1990s. Crystal Pepsi was a clear version of Pepsi introduced in 1992 to capitalize on the "clear craze," a marketing fad in which transparent versions of products were created to connote purity. Coca-Cola fought back by coming out with Tab Clear, a clear version of their diet cola Tab, which was considered inferior to the flagship cola products and reportedly meant to fail and sabotage Crystal Pepsi in the process. Tab Clear was marketed as a diet drink, a category that was less popular at the time. Because both drinks were clear colas, the negative perception of Tab Clear supposedly made consumers believe that Crystal Pepsi had similar undesirable qualities. As the former chief marketing officer of Coca-Cola asserted (Denny, 2013): "This is like a cola, but it doesn't have any color. It has all this great taste. And we said, 'No, Crystal Pepsi is actually a diet drink.' Even though it wasn't. Because Tab had the attributes of diet, which was its demise. That was its problem. It was perceived to be a medicinal drink. Within three or five months, Tab Clear was dead. And so was Crystal Pepsi." While the only record of this tactic came after the fact from the former executive (who stands to benefit from this narrative) and has not been otherwise corroborated to our knowledge, it usefully illustrates the conditions we proposed for the effect to emerge: Crystal Pepsi was a new product and so consumers were not sure how to feel about it, clear

colas formed an unfamiliar category so consumers were unsure about their qualities, and consumers plausibly felt that different clear colas had similar latent properties.

## General Discussion

Empirical demonstrations of context-dependent preferences have long been posed as challenges to traditional theories of economic rationality. And yet, the instability which so intrigues us is also what makes these phenomena hard to reliably grasp. Context effects still inspire decision scientists to search for their underlying principles, the stable mechanisms which can consistently explain when they will emerge and when they won't.

In this paper, we put forth an account of the repulsion effect, a recently documented context effect in which the presence of an inferior decoy option makes the target option which dominates it seem less appealing. It has been observed in both economic and perceptual judgment, and contrasts with the four-decade-old attraction effect. Our theory is based on the idea that the true values of options are only imperfectly signaled by their observable attributes, and thus people must draw upon all available cues to form evaluations. If the target and decoy are believed to have been generated by the same underlying process, the decoy can provide a negative signal about the value of the target. We constructed a hierarchical Bayesian model to formalize this pattern of reasoning. Our model embodies a point made by Simonson (2014): "a finding that a moldy orange taints an adjacent (yet nonmoldy) orange and generates repulsion suggests that certain inferior options infect similar options that are susceptible to the same affliction; in contrast, an overpriced, unattractive sweater has no bearing on the quality of a more attractive sweater and merely highlights its superior value. It thus seems that the nature of the relationship between the two adjacent options is one moderator of the resulting effect."

The only past attempts to formally model the repulsion effect have been restricted to sequential sampling models that lay out the dynamic process of decision making, and have so far proven insufficient to naturally capture the effect (Spektor et al., 2021, 2022). Our

normative approach offers a complementary perspective that casts light from a different angle. In the influential taxonomy of David Marr (1982), sequential sampling models reside on the "algorithmic" level which emphasizes the process by which a computation is achieved, while ours resides on the "computational" level which foregrounds the adaptive logic of the computation and abstracts away from the algorithm. No formal computational-level theory has been proposed before to explain the repulsion effect. Our goal was to fill this gap by developing such a theory and demonstrating its explanatory power. This perspective lets us view the problem in a way that brings clarity to many elements. It obscures other elements by necessity, especially aspects of the decision process such as presentation format and timing (Cataldo & Cohen, 2019, 2021a, 2021b), because computational-level accounts are inherently less suited for these than algorithmic ones.

The Bayesian framework nonetheless lets us flexibly encode many kinds of beliefs by altering priors and likelihoods at various levels of the hierarchy. This affords us a great deal of theoretical power, and allows exploration into the common and distinct properties of decision making across different markets, or across economic and perceptual domains. We made several simplifying assumptions to keep our exposition clean, and mentioned a few ways they could be profitably relaxed. Even more extensions may be possible.

Although our theory describes high-level computations rather than algorithmic processing, we conjecture that it could be transformed into a version which makes predictions about process variables. This would link it to an important strand of research which seeks to characterize contextual preference reversals using mechanistic models of evidence accumulation dynamics (e.g., Bhatia, 2013; Busemeyer, Gluth, Rieskamp, & Turner, 2019; Noguchi & Stewart, 2018; Roe et al., 2001; Spektor et al., 2018; Tsetsos, Usher, & Chater, 2010; Turner, Schley, Muller, & Tsetsos, 2018; Usher & McClelland, 2004). Models in this class are able to capture the joint distribution of choices and response times, and have more recently incorporated patterns of attention (e.g., Krajbich, 2019; Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011). They were originally

inspired by optimal statistical algorithms for hypothesis testing (Arrow, Blackwell, & Girshick, 1949; Wald & Wolfowitz, 1948) and can sometimes be expressed in Bayesian terms (Bitzer, Park, Blankenburg, & Kiebel, 2014; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Callaway, Rangel, & Griffiths, 2021; Fudenberg, Strack, & Strzalecki, 2018). Our model might be connected to this dynamic form if moment-to-moment evidence accumulation were based on noisy samples of fixated option values. Such specification of the decision making process could help predict the time course of repulsion (Spektor et al., 2018, 2022), similar to existing models of contextual deliberation (Guo, 2016, 2022).

Inference could also interact with other cognitive processes like memory (Kreps, 1990, p. 27). For example, options may not appear all at the same time, but rather may be be presented one after another. In this setting, repulsion effects have been observed when the target is presented first, and modeling of evidence accumulation reveals how memory decay is important in capturing the data (Evans, Holmes, Dasari, & Trueblood, 2021). Our explanation provides a complementary normative mechanism which could contribute to this phenomenon, in which memory decay is recast as noise in the retrieval process. When the target option is presented first, recollection of its attributes or their implied value will be especially noisy at the time of choice, and should be rationally biased toward the prior group mean which is informed by the decoy. This idea dovetails with other research demonstrating hierarchical memory encoding of perceptual stimuli following Bayesian principles (Brady & Alvarez, 2011; Hemmer & Steyvers, 2009a, 2009b; Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000), as well as Bayesian biases in memory-based evaluation of economic stimuli (Y. Li & Epley, 2009; Weilbächer, Kraemer, & Gluth, 2020). These kinds of links between levels of explanation can inspire new perspectives on context effects.

We focused on the repulsion effect in an effort to avoid retreading 40 years of historical debate over the attraction effect (Huber et al., 2014). The mechanism we propose is compatible with other sources of context effects. But there is no clear consensus on why

attraction effects are observed, even though there are many reasonable theories. Even the empirical boundaries of the attraction effect remain unclear, as work like that of Trendl et al. (2021) reveals. Therefore, any specific link that we might posit would be speculative. The fact that there are other mechanisms involved in decision making that could have other (even opposing) effects does not diminish our own. While integrating multiple components would be practically interesting for future work, it would obscure the mechanism we are focusing on here and presenting for the first time, making it harder to clearly grasp.

Nonetheless, we hope that our theory can contribute to the broader dialogue by helping to identify conditions under which the attraction effect may be opposed. Several have argued that ordinal information plays a key role in attraction (Howes et al., 2016; Natenzon, 2019) and that imprecise attribute representations can diminish it by obscuring the dominance relationship (e.g., Huber et al., 2014; Simonson, 2014; Spektor et al., 2021). This can explain why attraction tends to be observed when attributes are concrete and bear little ambiguity, such as price (Simonson, 2014). However, it has not been recognized that imprecision could also play a central role in the repulsion effect.

Furthermore, rather than drawing lines between qualitative and quantitative stimuli (Frederick et al., 2014; Yang & Lynn, 2014) or between perceptual and preferential stimuli (Spektor et al., 2018), our theory views them all through the common lens of inference and distinguishes attributes based on inferential uncertainty (Spektor et al., 2021). Although certain classes of attributes might be generally considered more precise (e.g., numeric versus verbal descriptions), variation can still exist within these classes. For instance, some attributes may be quantitative and yet imprecise, such as abstract quality ratings that are hard to interpret; others may be qualitative and yet precise, such as clear and detailed verbal or visual depictions. Much remains to be understood about how attribute perception depends on the format of presentation.

The existence of other Bayesian models of context effects offers hope that some aspects of these effects may be reconciled through common mechanisms. By virtue of the

Bayesian framework, our model could in principle be combined with others (e.g., Howes et al., 2016; Shenoy & Yu, 2013) into a mega-model that includes multiple kinds of uncertainty and makes predictions about what will happen when all forces are considered in sum. This prospect also highlights the care needed when specifying uncertainty, as imprecision along some dimensions can lead to attraction while others foster repulsion. Although a full account of context dependence will surely need to incorporate other cognitive mechanisms, we believe the elements we have laid out here provide a valuable piece of the puzzle.

References

Acuña, D. E., & Schrater, P. (2010). Structure learning in human sequential decision-making. *PLOS Computational Biology*, *6*(12), e1001003.

Ahmad, S., & Yu, A. J. (2015). A rational model for individual differences in preference choice. In *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 54–59).

Aldous, D. J. (1985). Exchangeability and related topics. In *École d'été de probabilités de saint-flour xiii—1983* (pp. 1–198). Springer.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409.

André, Q., Reinholtz, N., & De Langhe, B. (2021). Can consumers learn price dispersion? evidence for dispersion spillover across categories. *Journal of Consumer Research*.

Ariely, D. (2008). *Predictably irrational*. HarperCollins.

Arrow, K. J., Blackwell, D., & Girshick, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica*, 213–244.

Austerweil, J. L., Gershman, S. J., Tenenbaum, J. B., & Griffiths, T. L. (2015). Structure and flexibility in Bayesian models of cognition. In *Oxford Handbook of Computational and Mathematical Psychology* (pp. 187–208). Oxford University Press Oxford, UK.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer.

Berkowitsch, N. A., Scheibehenne, B., Rieskamp, J., & Matthäus, M. (2015). A generalized distance function for preferential choices. *British Journal of Mathematical and Statistical Psychology*, *68*(2), 310–325.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*(3), 522.

Bhui, R. (2018). Case-based decision neuroscience: Economic judgment by similarity. In *Goal-directed decision making: Computations and neural circuits* (pp. 67–103). Elsevier.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

Bill, J., Pailian, H., Gershman, S. J., & Drugowitsch, J. (2020). Hierarchical structure is employed by humans during visual motion perception. *Proceedings of the National Academy of Sciences*, *117*(39), 24581–24589.

Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, *8*, 102.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, *121*(5), 803–843.

Bordley, R. F. (1992). An intransitive expectations-based Bayesian variant of prospect theory. *Journal of Risk and Uncertainty*, *5*(2), 127–144.

Bottomley, P. A., & Holden, S. J. S. (2001). Do we really know how consumers evaluate brand extensions? empirical generalizations based on secondary analysis of eight studies. *Journal of Marketing Research*, *38*(4), 494–500.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Brendl, M. C., Atasoy, Ö., & Samson, C. (in press). Preferential attraction effects with visual stimuli: The role of quantitative versus qualitative visual attributes. *Psychological Science*.

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences*, *23*(3), 251–263.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology*, *17*(3), e1008863.

Cao, Y., Summerfield, C., Park, H., Giordano, B. L., & Kayser, C. (2019). Causal inference in the multisensory brain. *Neuron*, *102*(5), 1076–1087.

Cataldo, A. M., & Cohen, A. L. (2019). The comparison process as an account of variation in the attraction, compromise, and similarity effects. *Psychonomic Bulletin & Review*, *26*(3), 934–942.

Cataldo, A. M., & Cohen, A. L. (2021a). The influence of within-alternative and within-dimension similarity on context effects. *Decision*, *8*(3), 202.

Cataldo, A. M., & Cohen, A. L. (2021b). Modeling preference reversals in context effects over time. *Computational Brain & Behavior*, *4*(1), 101–123.

Ching, A. T., Erdem, T., & Keane, M. P. (2013). Learning models: An assessment of progress, challenges, and new developments. *Marketing Science*, *32*(6), 913–938.

Daviet, R., & Webb, R. (2020). A double decoy experiment to distinguish theories of dominance effects. *Available at SSRN 3374514*.

Denny, S. (2013). *Killing giants: 10 strategies to topple the goliath in your industry*. Penguin.

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., ... Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLOS Computational Biology*, *10*(9), e1003810.

Dorfman, H. M., Bhui, R., Hughes, B. L., & Gershman, S. J. (2019). Causal inference about good and bad outcomes. *Psychological Science*, *30*(4), 516–525.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

Erdem, T. (1998). An empirical analysis of umbrella branding. *Journal of Marketing Research*, *35*(3), 339–351.

Erdem, T., & Chang, S. R. (2012). A cross-category and cross-country analysis of umbrella branding for national and store brands. *Journal of the Academy of Marketing Science*, *40*(1), 86–101.

Erdem, T., & Sun, B. (2002). An empirical investigation of the spillover effects of advertising and sales promotions in umbrella branding. *Journal of Marketing Research*, *39*(4), 408–420.

Erdem, T., & Swait, J. (1998). Brand equity as a signaling. *Journal of Consumer Psychology*, *7*(2), 131–157.

Erdem, T., Swait, J., & Valenzuela, A. (2006). Brands as signals: A cross-country validation study. *Journal of Marketing*, *70*(1), 34–49.

Ert, E., & Lejarraga, T. (2018). The effect of experience on context-dependent decisions. *Journal of Behavioral Decision Making*, *31*(4), 535–546.

Evangelidis, I., Levav, J., & Simonson, I. (2018). The asymmetric impact of context on advantaged versus disadvantaged options. *Journal of Marketing Research*, *55*(2), 239–253.

Evans, N. J., Holmes, W. R., Dasari, A., & Trueblood, J. S. (2021). The impact of presentation order on attraction and repulsion effects in decision-making. *Decision*, *8*(1), 36.

Farquhar, P. H., & Pratkanis, A. R. (1993). Decision structuring with phantom alternatives. *Management Science*, *39*(10), 1214–1226.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117–141). Cambridge University Press.

Fox, C. R., & Hadar, L. (2006). "Decisions from experience" = sampling error + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*(2), 159.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing*

*Research*, *51*(4), 487–507.

Fudenberg, D., Strack, P., & Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, *108*(12), 3651–84.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12.

Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, *117*(1), 197.

Gershman, S. J., & Cikara, M. (2020). Social-structure learning. *Current Directions in Psychological Science*, *29*(5), 460–466.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, *5*, 43–50.

Gershman, S. J., Pouncy, H. T., & Gweon, H. (2017). Learning the structure of social influence. *Cognitive Science*, *41*, 545–575.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge Handbook of Computational Cognitive Modeling.* Carnegie Mellon University.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, *116*(4), 661.

Guo, L. (2016). Contextual deliberation and preference construction. *Management*

*Science*, *62*(10), 2977–2993.

Guo, L. (2022). Testing the role of contextual deliberation in the compromise effect. *Management Science*, *68*(6), 4326–4355.

Hardt, O., & Pohl, R. (2003). Hindsight bias as a function of anchor distance and anchor plausibility. *Memory*, *11*(4-5), 379–394.

Hedgcock, W., & Rao, A. R. (2009). Trade-off aversion as an explanation for the attraction effect: A functional magnetic resonance imaging study. *Journal of Marketing Research*, *46*(1), 1–13.

Hemmer, P., & Steyvers, M. (2009a). A Bayesian account of reconstructive memory. *Topics in Cognitive Science*, *1*(1), 189–202.

Hemmer, P., & Steyvers, M. (2009b). Integrating episodic memories and prior knowledge at multiple levels of abstraction. *Psychonomic Bulletin & Review*, *16*(1), 80–87.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523.

Highhouse, S. (1996). Context-dependent selection: The effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes*, *65*(1), 68–76.

Howes, A., Warren, P. A., Farmer, G., El-Deredy, W., & Lewis, R. L. (2016). Why contextual preference reversals maximize expected value. *Psychological Review*, *123*(4), 368.

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

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, *51*(4), 520–525.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars:

Prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General*, *129*(2), 220.

Izakson, L., Zeevi, Y., & Levy, D. J. (2020). Attraction to similar options: The gestalt law of proximity is related to the attraction effect. *PLOS One*, *15*(10), e0240937.

Kamenica, E. (2008). Contextual inference in markets: On the informational content of product lines. *American Economic Review*, *98*(5), 2127–49.

Kardes, F. R., Posavac, S. S., Cronley, M. L., & Herr, P. M. (2008). Consumer inference. In *Handbook of consumer psychology* (pp. 165–192). Taylor & Francis.

Keller, K. L. (2012). Economic and behavioral perspectives on brand extension. *Marketing Science*, 772–776.

Kemp, C., Bernstein, A., & Tenenbaum, J. B. (2005). A generative theory of similarity. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1132–1137).

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS One*, *2*(9), e943.

Krajbich, I. (2019). Accounting for attention in sequential sampling models of decision making. *Current Opinion in Psychology*, *29*, 6–11.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Kreps, D. M. (1990). *A course in microeconomic theory*. Princeton University Press.

Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *eLife*, *9*, e53162.

Lau, T., Pouncy, H. T., Gershman, S. J., & Cikara, M. (2018). Discovering social groups

via latent structure learning. *Journal of Experimental Psychology: General*, *147*(12), 1881.

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

Levin, I. P., & Levin, A. M. (2000). Modeling the role of brand alliances in the assimilation of product evaluations. *Journal of Consumer Psychology*, *9*(1), 43–52.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311.

Li, S., & Yu, N. N. (2018). Context-dependent choice as explained by foraging theory. *Journal of Economic Theory*, *175*, 159–177.

Li, V., Michael, E., Balaguer, J., Castañón, S. H., & Summerfield, C. (2018). Gain control explains the effect of distraction in human perceptual, cognitive, and economic decision making. *Proceedings of the National Academy of Sciences*, *115*(38), E8825–E8834.

Li, Y., & Epley, N. (2009). When the best appears to be saved for last: Serial position effects on choice. *Journal of Behavioral Decision Making*, *22*(4), 378–389.

Liao, J., Chen, Y., Lin, W., & Mo, L. (2020). The influence of distance between decoy and target on context effect: Attraction or repulsion? *Journal of Behavioral Decision Making*, *34*(3), 432–447.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*.

Loken, B., Barsalou, L. W., & Joiner, C. (2008). Categorization theory and research in consumer psychology: Category representation and category-based inference. In *Handbook of consumer psychology* (pp. 133–164). Taylor & Francis.

Luce, R. D. (1959). *Individual choice behavior*. New York: John Wiley.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. Dover Publications.

Marcoul, P., & Weninger, Q. (2008). Search and active learning with correlated information: Empirical evidence from mid-atlantic clam fishermen. *Journal of Economic Dynamics and Control*, *32*(6), 1921–1948.

Markman, A. B., & Medin, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes*, *63*(2), 117–130.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39.

Mathys, C., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience*, *8*, 825.

McKenzie, C. R., Sher, S., Leong, L. M., & Müller-Trede, J. (2018). Constructed preferences, rationality, and choice architecture. *Review of Behavioral Economics*, *5*(3-4), 337–360.

Miklós-Thal, J. (2012). Linking reputations through umbrella branding. *Quantitative Marketing and Economics*, *10*(3), 335–374.

Mishra, S., Umesh, U., & Stem, D. E., Jr. (1993). Antecedents of the attraction effect: An information-processing approach. *Journal of Marketing Research*, *30*(3), 331–349.

Moorthy, S. (2012). Can brand extension signal product quality? *Marketing Science*, *31*(5), 756–770.

Natenzon, P. (2019). Random choice and learning. *Journal of Political Economy*, *127*(1), 419–457.

Navarro, D. J., & Kemp, C. (2017). None of the above: A Bayesian account of the detection of novel categories. *Psychological Review*, *124*(5), 643.

Noguchi, T., & Stewart, N. (2018). Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological Review*, *125*(4), 512.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Orhun, A. Y. (2009). Optimal product line design when consumers exhibit choice set-dependent preferences. *Marketing Science*, *28*(5), 868–886.

Pettibone, J. C., & Wedell, D. H. (2000). Examining models of nondominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*(2), 300–328.

Pettibone, J. C., & Wedell, D. H. (2007). Testing alternative explanations of phantom decoy effects. *Journal of Behavioral Decision Making*, *20*(3), 323–341.

Pratkanis, A. R., & Farquhar, P. H. (1992). A brief history of research on phantom alternatives: Evidence for seven empirical generalizations about phantoms. *Basic and Applied Social Psychology*, *13*(1), 103–122.

Ratneshwar, S., Shocker, A. D., & Stewart, D. W. (1987). Toward understanding the attraction effect: The implications of product stimulus meaningfulness and familiarity. *Journal of Consumer Research*, *13*(4), 520–533.

Rigoli, F., Mathys, C., Friston, K. J., & Dolan, R. J. (2017). A unifying Bayesian account of contextual effects in value-based choice. *PLOS Computational Biology*, *13*(10), e1005769.

Roberts, J. H., & Urban, G. L. (1988). Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science*, *34*(2), 167–185.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological Review*, *108*(2), 370.

Rooderkerk, R. P., Van Heerde, H. J., & Bijmolt, T. H. A. (2011). Incorporating context

effects into a choice model. *Journal of Marketing Research*, *48*(4), 767–780.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation*, *19*(12), 3335–3355.

Scarpi, D. (2011). The impact of phantom decoys on choices in cats. *Animal Cognition*, *14*(1), 127–136.

Scarpi, D., & Pizzi, G. (2013). The impact of phantom decoys on choices and perceptions. *Journal of Behavioral Decision Making*, *26*(5), 451–461.

Schulz, E., Franklin, N. T., & Gershman, S. J. (2020). Finding structure in multi-armed bandits. *Cognitive Psychology*, *119*, 101261.

Sen, S. (1998). Knowledge, information mode, and the attraction effect. *Journal of Consumer Research*, *25*(1), 64–77.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, *14*(9), 425–432.

Sharp, P. B., Fradkin, I., & Eldar, E. (2022). Hierarchical inference as a source of human biases. *Cognitive, Affective, & Behavioral Neuroscience*, 1–15.

Shenoy, P., & Yu, A. J. (2013). Rational preference shifts in multi-attribute choice: What is fair? In *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 1300–1305).

Sher, S., & McKenzie, C. R. M. (2014). Options as information: Rational reversals of evaluation and preference. *Journal of Experimental Psychology: General*, *143*(3), 1127.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174.

Simonson, I. (2014). Vices and virtues of misguided replications: The case of asymmetric dominance. *Journal of Marketing Research*, *51*(4), 514–519.

Soltani, A., De Martino, B., & Camerer, C. (2012). A range-normalization model of

context-dependent choice: A new model and evidence. *PLOS Computational Biology*, *8*(7), e1002607.

Spektor, M. S., Bhatia, S., & Gluth, S. (2021). The elusiveness of context effects in decision making. *Trends in Cognitive Sciences*.

Spektor, M. S., Gluth, S., Fontanesi, L., & Rieskamp, J. (2019). How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological Review*, *126*(1), 52.

Spektor, M. S., Kellen, D., & Hotaling, J. M. (2018). When the good looks bad: An experimental exploration of the repulsion effect. *Psychological Science*, *29*(8), 1309–1320.

Spektor, M. S., Kellen, D., & Klauer, K. C. (2022). The repulsion effect in preferential choice and its relation to perceptual choice. *Cognition*, *225*, 105164.

Sridhar, K., Bezawada, R., & Trivedi, M. (2012). Investigating the drivers of consumer cross-category learning for new products using multiple data sets. *Marketing Science*, *31*(4), 668–688.

Sullivan, M. (1990). Measuring image spillovers in umbrella-branded products. *Journal of Business*, 309–329.

Summerfield, C., & Parpart, P. (2022). Normative principles for decision-making in natural environments. *Annual Review of Psychology*, *73*, 53–77.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Trendl, A., Stewart, N., & Mullett, T. L. (2021). A zero attraction effect in naturalistic choice. *Decision*, *8*(1), 55.

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*(4), 1275.

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, *6*(1), 1–12.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*(3), 371.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179–1189.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*(3), 757.

Viscusi, W. K. (1989). Prospective reference theory: Toward an explanation of the paradoxes. *Journal of Risk and Uncertainty*, *2*(3), 235–263.

Volkmann, J. (1951). Scales of judgment and their implications for social psychology. In *Social psychology at the crossroads; The University of Oklahoma lectures in social psychology* (pp. 273–298). Harper.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 326–339.

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

Weilbächer, R. A., Kraemer, P. M., & Gluth, S. (2020). The reflection effect in memory-based decisions. *Psychological Science*, *31*(11), 1439–1451.

Wernerfelt, B. (1988). Umbrella branding as a signal of new product quality: An example

of signalling by posting a bond. *RAND Journal of Economics*, 458–466.

Wernerfelt, B. (1995). A rational reconstruction of the compromise effect: Using market data to infer utilities. *Journal of Consumer Research*, *21*(4), 627–633.

Wernerfelt, B. (2012). On brand extension as a signal of product quality. *Marketing Science*.

Wilson, S. A., Arora, S., Zhang, Q., & Griffiths, T. L. (2021). A rational account of anchor effects in hindsight bias. In *Proceedings of the 43rd annual meeting of the cognitive science society.*

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, *12*, 579–601.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature Human Behaviour*, *2*(12), 915–924.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245.

Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research*, *51*(4), 508–513.