

# Pairwise Normalization: A Neuroeconomic Theory of Multi-Attribute Choice\*

Peter Landry<sup>†</sup>

Ryan Webb<sup>‡</sup>

April 18, 2019

## Abstract

We present a theory of multi-attribute choice founded in the neuroscience of perception. According to our theory, valuation is formed through a series of pairwise, attribute-level comparisons implemented by *(divisive) normalization* — a normatively-grounded form of relative value coding observed across sensory modalities and in species ranging from honeybees to humans. As we demonstrate, “pairwise normalization” captures a broad range of behavioral regularities, including the compromise and asymmetric dominance effects, the diversification bias in allocation decisions, and majority-rule preference cycles (among several others). In binary choice, the model also offers a potential neurobiological foundation for Cobb-Douglas preferences and other classic microeconomic preference representations.

---

\* We thank Pedro Bordalo, David Freeman, Kenway Louie, Antonio Rangel, Colin Stewart, and especially Anthony Marley for many helpful comments on this paper. We also received valuable feedback from seminar participants at Stanford Graduate School of Business, California Institute of Technology, McGill, Carnegie-Mellon, Simon Fraser University, Maastricht University, University of Zurich, Virginia Tech, the Barcelona Summer Forum, BRIC 2018, and the Sloan-Nomis Workshop. Fiorella Wever provided excellent research assistance. We also thank Alireza Soltani, Benedetto de Martino, and Colin Camerer for providing experimental data.

<sup>†</sup> Rotman School of Management, University of Toronto, and Department of Management, University of Toronto, Mississauga. E-mail: peter.landry@rotman.utoronto.ca.

<sup>‡</sup> Rotman School of Management, University of Toronto, and Department of Economics, University of Toronto. E-mail: ryan.webb@rotman.utoronto.ca.

# 1 Introduction

Standard choice theories presume that an individual’s valuation of an alternative does not depend on the set of alternatives under consideration. However, a large empirical literature has revealed several violations of such “context-independence.” For example, simply adding an alternative to a choice set can alter preferences among existing alternatives (see Rieskamp et al., 2006, for a review). Empirical demonstrations of context effects can be found in both laboratory experiments (beginning with Huber et al., 1982, and Simonson, 1989) and in field data (e.g. Doyle et al., 1999; Geyskens et al., 2010), and extend to many types of decisions — including consumer choice, choices among lotteries, doctors’ prescription decisions, perceptual decisions, and mate selection, to name just a few.<sup>1</sup>

Though less familiar to behavioral researchers, context-independence is also challenged by an established neuroscience literature (beginning with Hartline and Ratliff, 1957) demonstrating that the brain encodes information in relative, not absolute terms. For example, the neural activity encoding the value of an alternative decreases — indicating a reduced valuation — as the value of another alternative rises (Louie et al., 2011; Holper et al., 2017). This pattern of neural activity is consistent with *(divisive) normalization* — a well-documented and normatively-grounded neural computation originally used to model the mechanisms of visual perception and more recently applied to value-based choice (see Rangel and Clithero, 2012, Carandini and Heeger, 2012, and Louie et al., 2015, for reviews).

In its simplest conceivable form, the normalization computation merely re-expresses some input value  $a$  — which may represent the utility of an alternative, or the intensity of a sensory stimuli (such as the brightness of a pixel) — relative to another input  $b$  as  $\frac{a}{a+b}$ . Indeed, the prevailing neuroscience literature conceptualizes such “division by neurons” as an arithmetic operation that is *actually performed* in the brain.<sup>2</sup>

Why wouldn’t our brains just encode  $a$  independently of  $b$ ? The answer is thought to stem from biological constraints. The brain has a limited number of neurons, each with a bounded response range. Thus, information must be compressed within these bounds. A relative value encoding is then needed to ensure this compression is well-calibrated to the choice environment (a point noted in the economics literature by Robson, 2001, and Rayo and Becker, 2007; also see Netzer, 2009, Woodford, 2012, and Robson and Whitehead, 2018). A relative encoding using the normalization computation has been shown to optimally mitigate choice mistakes subject to these biological constraints (Webb

---

<sup>1</sup>See, for example, Huber et al. (2014), Soltani et al. (2012), Schwartz and Chapman (1999), Trueblood et al. (2013), Lea and Ryan (2015).

<sup>2</sup>See, for example, Carandini and Heeger’s (1994) article in *Science* (which coined the phrase in quotes) as well as Wilson et al.’s (2012) closely-related work in *Nature*. As shown by Louie et al. (2014), this divisive functional form can be derived as the equilibrium solution to the system of differential equations that govern neural activity in a stylized neural circuit.

et al., 2016; Steverson et al., 2017), as normalization efficiently facilitates the perception of both large and small differences on a common scale — e.g. helping to distinguish “one dollar from two dollars and one million dollars from two million dollars” (Carandini and Heeger, 2012).

In this paper, we explore whether this inherently context-dependent computation might relate to context-dependent behavior. To do so, we adapt the “ $\frac{a}{a+b}$ ” normalization model to the setting where behavioral research on context-dependence is mainly focused: multi-attribute choice. Specifically, if  $\mathbf{x}$  is an alternative with  $x_1, \dots, x_N$  denoting its  $N$  attribute values, the decision-maker’s valuation of  $\mathbf{x}$  according to our basic *pairwise normalization* (PN) model is normalized relative to other alternatives in the choice set  $X$  as:

$$V(\mathbf{x}; X) = \sum_{n=1}^N \sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{x_n}{x_n + y_n}.$$

This formulation is “pairwise” in the sense that each term reflects an attribute-level comparison (normalization) of  $\mathbf{x}$  to some other alternative  $\mathbf{y}$ . Pairwise comparisons have long been a feature of multi-attribute choice theories (e.g. Tversky and Simonson, 1993) and have substantial empirical support from eye-tracking studies showing that individuals typically compare multi-attribute alternatives in pairs on one attribute dimension at a time. In context-dependent choice, in particular, Noguchi and Stewart (2014) find that “alternatives are repeatedly compared in pairs on single dimensions.”<sup>3</sup>

Our modelling approach demonstrates how neuroscience may prove useful to economists as a source of candidate functional form representations to consider in model selection (as suggested by Bernheim, 2009). Arguably the simplest, standard multi-attribute choice model is an additive model,  $V(\mathbf{x}) = \sum_n x_n$ . This additive model provides a common foundation for many leading multi-attribute choice theories that address context-dependence (e.g. Tversky and Simonson, 1993; Kivetz et al. 2004a; Koszegi and Szeidl, 2013; Bordalo et al. 2013; and Bushong et al. 2017).<sup>4</sup> These theories typically replace each term in the summation with a function of  $x_n$  that also depends on the set of alternatives. Similarly, our theory modifies the additive model by applying pairwise normalization to each attribute value (effectively replacing  $x_n$  with  $\sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{x_n}{x_n + y_n}$ ). This formalizes pairwise normalization in its most elemental form, isolated from other factors that may influence choice, and with minimal parametric freedom. Despite its simplicity, the model’s predictions capture

<sup>3</sup> Also see Russo and Doshier (1983), Arieli et al. (2011), Noguchi and Stewart (2018), as well as Russo and Rosen (1975), who emphasize that the predominance of pairwise comparisons in choice may stem from cognitive constraints, as even ternary comparisons (which they observed roughly 2 percent as often as pairwise comparisons) can stretch working memory to its limits. For a lengthier discussion of pairwise comparisons in relation to other theoretical representations of attribute-level comparisons, see Section 8.

<sup>4</sup> We consider these relationships more deeply in Sections 7 and 8 (see, in particular, Figure 4 and Table 2).

a broad range of context-dependent behavioral regularities. See Table 1.<sup>5</sup>

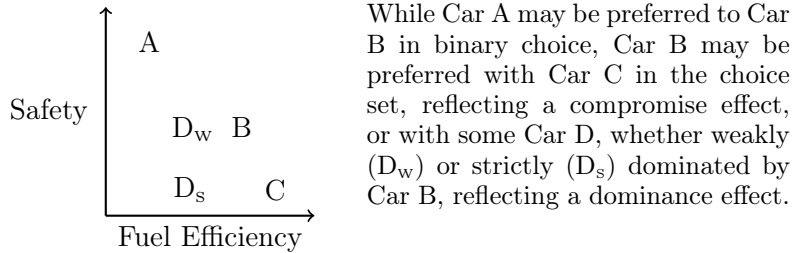
The rest of this paper proceeds as follows. Section 2 reviews the behavioral regularities listed in Table 1. Section 3 presents the theoretical model. Section 4 examines how a preference between two alternatives can be affected by a third alternative, and relates these effects to the notion advanced by Tversky and Russo (1969) and Natenzon (2018) that similar alternatives are “easy to compare.” Section 5 considers choices among alternatives defined on three dimensions. Section 6 considers various allocation problems. Section 7 explores a one-parameter generalization of the model. Section 8 elaborates on the varying representations of attribute-level comparisons in the relevant theoretical literature.

## 2 Behavioral Patterns: A Review

We now review the behavioral patterns listed in Table 1.

(I) *Compromise Effect*. The ‘compromise effect’ refers to the tendency for decision-makers — whether subjects in a laboratory experiment (e.g. Simonson, 1989) or real-world shoppers (Geyskens et al., 2010) — to show a stronger preference for an alternative if it is presented in a choice set where it is the intermediate option on each dimension. For example, if car A is safer but less fuel-efficient than car B, an individual who prefers A to B in a binary choice may instead prefer B when a third car is included, C, that is even less safe and more efficient than B (see Figure 1).

Figure 1. Illustration of Compromise and Dominance Effects



(II) *Dominance Effect*. The ‘(asymmetric) dominance effect,’ also known as the ‘attraction’ or ‘decoy’ effect, refers to the tendency to show a stronger preference for an alternative when presented with a ‘decoy’ that is worse on each dimension (e.g. Huber et al., 1982; Doyle et al., 1999). That is, while the safer but less efficient car A is preferred to car B in a binary choice, B may be preferred with the addition of a decoy, car D, that is even less safe and less efficient than B (yet still more efficient than A). Though sometimes demonstrated using weakly-dominated decoys that match the dominant alternative on its comparatively weak attribute dimension (e.g. Kivetz et al.’s, 2004b, economist subscription study), the dominance effect appears stronger for decoys that are worse — thereby

<sup>5</sup>For detailed explanations of how each model’s predictions were classified in Table 1, see Appendix D.

Table 1. Behavioral Patterns Generated by Pairwise Normalization\*

		Pairwise Normalization	Bordalo, Gennaioli, Shleifer (2013)	Bushong, Rabin, Schwartzstein (2017)	Kivetz, Netzer, Srinivasan (2004a)	Tversky & Simonson (1993)
(I)	Compromise Effect	<b>Y</b>	S	<b>Y</b>	<b>Y</b>	<b>Y</b>
(II)	Dominance Effect $\left\{ \begin{array}{l} \text{Weak} \\ \text{Strict} \end{array} \right.$	<b>Y</b>	S	N	N	<b>Y</b>
		<b>Y</b>	S	<b>Y</b>	<b>Y</b>	<b>Y</b>
	...Decoy-Range Effect	<b>Y</b>	S	<b>Y</b>	<b>Y</b>	N
(III)	Relative Difference Effect	<b>Y</b>	S	N	N	N
(IV)	Majority-Rule Preference Cycles	<b>Y</b>	S	<b>Y</b>	<b>Y</b>	N
(V)	Splitting Bias	<b>Y</b>	S	<b>Y</b>	<b>Y</b>	N
(VI)	Alignability Effect	<b>Y</b>	S	<b>Y</b>	<b>Y</b>	N
(VII)	Diversification Bias	<b>Y</b>	S	N	N	N
(VIII)	Feature Bias	<b>Y</b>	<b>Y</b>	N	N	N

Key behavioral regularities predicted by the basic PN model as compared to several prevailing multi-attribute choice theories. Here, ‘**Y**’ means the model robustly predicts the behavior (i.e. never predicts the opposite or no effect under conditions for which it would be expected), ‘N’ means the model does not predict the behavior, and ‘S’ means the model sometimes predicts the behavior and sometimes predicts the opposite effect. See Appendix D for a detailed explanation of how these predictions were classified and Figure 2 for illustrations of the predictions for items (I) and (II).

\* This table only includes theories that are directly comparable to the basic PN model in that the domains of their analyses have sufficient overlap with ours. For instance, to consistently classify a theory’s predictions for the above items, alternatives must be defined on exogenous attribute dimensions and the theory must allow more than two attributes. Notable theories addressing the compromise and/or dominance effects in somewhat different domains include Kamenica’s (2008) contextual inference theory (which, unlike the theories listed above, models a market with both consumers and a firm), de Clippel and Eliaz’s (2012) dual-self intrapersonal bargaining theory, Soltani et al.’s (2012) theory of range (instead of divisive) normalization across two attributes, Ok et al.’s (2015) endogenous reference point theory, and Natenzon’s (2018) Bayesian probit theory. That said, a static version of Koszegi and Szeidl’s (2013) dynamic ‘focusing’ model is directly comparable to our model. While this theory does not seek to address the context-dependent phenomena addressed by other theories, for completeness we derive their predictions for the above items in Appendix D.

expanding the range between the best and worst alternatives — on that dimension (Huber et al., 1982; Soltani et al., 2012). This ‘decoy-range effect’ suggests that a preference reversal from car A to car B is more likely with the strictly dominated decoy  $D_s$  than with the weakly dominated  $D_w$ .

(III) *Relative Difference Effect*. The ‘relative difference effect’ refers to the tendency to treat a difference between small values as if it were greater than an equal-sized difference between large values. For example, Kahneman and Tversky (1984) find that people are often willing to drive twenty minutes to save \$5 on a \$15 calculator, but not to save \$5 on a \$125 jacket — a finding that has since been confirmed and generalized by many others (see Azar, 2008, for a review). Among many illustrations of relative difference effects in the contingent valuation literature, Shiell and Gold (2002) find that subjects value immunity to a syndrome more on its own than as part of a bundle that already includes immunity to another syndrome.

(IV) *Majority-Rule Preference Cycles*. Suppose each of three potential alternatives is best on one dimension, second best on another, and worst on a third as follows:

	Attribute 1	Attribute 2	Attribute 3
Alternative A	Best	Middle	Worst
Alternative B	Middle	Worst	Best
Alternative C	Worst	Best	Middle

As shown by May (1954), binary choices among three such alternatives often exhibit a ‘majority-rule preference cycle’ whereby each alternative is preferred to that for which it is better on two of three attributes — here, A would be preferred to B, B would be preferred to C, yet C would be preferred to A.<sup>6</sup> In a recent study with alternatives designed such to put subjects on the cusp of indifference, Tsetsos et al. (2016) show that majority-rule preference cycles can even be more common than transitive preferences.

(V) *Splitting Bias*. The ‘splitting bias’ refers to the tendency to place more (cumulative) weight on an attribute when it is split into two subattributes. For example, job applicants weighted “job security” of a potential job more heavily if the attribute was decomposed into “personal job security” and “stability of the firm/risk of bankruptcy,” and likewise weighted “income” more heavily if it was decomposed into “starting salary” and “future salary increases” (Weber et al., 1988).<sup>7</sup> Two direct analogs (or arguably special cases)

<sup>6</sup>In May’s experiment, 17 of 62 subjects exhibited this particular preference ordering among hypothetical spouses, while no subjects exhibited the opposite ‘minority-rule’ cycle. As detailed in Appendix B, we also conducted an incentivized experiment confirming a statistically significant tendency for subjects to exhibit majority-rule preference cycles in choices among vacation packages. Russo and Doshier (1983) and Zhang et al. (2006) similarly report a tendency for subjects to overvalue an alternative in relation to another for which it is better on most attribute dimensions.

<sup>7</sup>Also see Weber and Borcherting (1993) for a brief review of this literature, as well as more recent evidence by Jacobi and Hobbs (2007) and Hamalainen and Alaja (2008).

of the splitting bias are the ‘event-splitting’ (or ‘coalescing’) effect in risky choice, which refers to the tendency to value a probabilistic reward more if the event for which the reward is attained is described as two sub-events (Starmer and Sugden, 1993; Humphrey, 1995; Birnbaum and Bahra, 2007), and the ‘part-whole bias’ in contingent valuation, which refers to the tendency to value a good more when its components are evaluated separately than when evaluated holistically (Kahneman and Knetsch, 1992; Bateman et al., 1997).

(VI) *Alignability Effect*. The ‘alignability effect’ refers to the tendency to place more weight on an attribute that is ‘alignable’ in the sense that it is present (though not necessarily equal) for all alternatives (Markman and Medin, 1995; Zhang and Markman, 1998; Gourville and Soman, 2005). For example, when considering a 1000-watt microwave or a 1100-watt microwave, one of which has a moisture sensor and the other an adjustable-speed turntable, the alignability effect implies that the wattage difference may be overweighted relative to the other, nonalignable features. Similarly, individuals tend to weight alignable attributes more heavily when alternatives are evaluated jointly rather than separately. For example, Hsee et al. (1999) find that a complete 24-piece dinnerware set is often rated more favorably than an incomplete 31-piece set when the sets are separately rated, but not when they are jointly rated.

(VII) *Diversification Bias*. The ‘diversification bias’ refers to the tendency to disproportionately favor equal allocations of an asset or resource across its components. For instance, Read and Loewenstein (1995) find that Halloween trick-or-treaters often select a mixed bundle of candy bars featuring one Milky Way and one Three Musketeers over a bundle with two of the same kind, despite selecting the same candy bar in two consecutive choices between one Milky Way and one Three Musketeers. In quite different settings, diversified gambles are often preferred to undiversified gambles that stochastically dominate the former (Rubinstein, 2002), while investors often exhibit a strong preference for savings plans that allocate contributions equally across the different funds included in the plan (Benartzi and Thaler, 2001; Bardolet et al., 2007).<sup>8</sup>

(VIII) *Feature Bias*. The ‘(extra) feature bias’ refers to the tendency to overvalue products with the most available features. As one example, demand for a video game rises substantially after the development of a new “button” or “scrollbar” control, despite buyers’ negligible usage of the new feature (Meyer et al., 2008), while more generally buyers commonly report dissatisfaction, stress, and anxiety with many-feature products after purchase (Thompson et al., 2005; Mick and Fournier, 1998). For a firm, the addition of an irrelevant product feature may create a sustained competitive advantage, even when consumers acknowledge the feature’s irrelevance (Carpenter et al., 1994). Supply-side

---

<sup>8</sup> As noted in these studies, investors favoring equal allocations across funds will end up investing more (or less) in stocks than in bonds simply because the available plans happen to include a greater (smaller) number of stock funds than bond funds.

responses to the feature bias appear to be common in light of the widely-noted proliferation of products with an excessive number of features — a trend known as “feature bloat” or “feature creep” (Thompson and Norton, 2011).

### 3 Basic Theoretical Model

Our basic setting is standard. A decision-maker (DM) faces a choice set  $X$ , where each  $\mathbf{x} = (x_1, \dots, x_N) \in X$  is defined on  $N > 0$  attribute dimensions and  $x_n \geq 0$  denotes  $\mathbf{x}$ ’s unnormalized attribute value (in utils) on attribute  $n$ .<sup>9</sup> In the basic PN model, the DM’s valuation of  $\mathbf{x}$  is given by:

$$V(\mathbf{x}; X) = \sum_{n=1}^N \sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{x_n}{x_n + y_n}, \quad (1)$$

where the DM (strictly) prefers  $\mathbf{x}$  to  $\mathbf{y}$  given  $X$  if  $V(\mathbf{x}; X) > V(\mathbf{y}; X)$  and is indifferent if  $V(\mathbf{x}; X) = V(\mathbf{y}; X)$ . Although  $\frac{x_n}{x_n + y_n}$  is undefined when  $x_n = y_n = 0$ , this case will not be relevant to our analysis. For this reason and without loss of insight, we assume throughout that, for all  $n \leq N$ , there is at most one  $\mathbf{x} \in X$  with  $x_n = 0$ .

The normalized valuation in (1) can be thought of as arising from a series of pairwise comparisons, where each of  $\mathbf{x}$ ’s attribute values are normalized in relation to the corresponding attribute value of each other alternative  $\mathbf{y} \in X \setminus \mathbf{x}$ . That is, when ‘compared’ to  $\mathbf{y}$ , the normalized value of  $\mathbf{x}$  on attribute  $n$  is simply  $\frac{x_n}{x_n + y_n}$ , while the overall valuation of  $\mathbf{x}$  is the sum of all such terms.<sup>10,11</sup> Note, in the baseline model it is implicit that the DM attends to all attributes of all alternatives when computing  $V(\mathbf{x}; X)$ .<sup>12</sup> A setting in which there are too many attributes and/or alternatives can be readily accounted for if  $X$  and  $n = 1, \dots, N$  instead represent the subsets of alternatives and attributes that are attended to, as in Noguchi and Stewart (2018). To see how our theoretical model may be adapted for empirical analyses, including subjective weights on measurable attributes defined in non-util units, see Daviet and Webb (2019). Also see Sullivan et al. (2019) for

<sup>9</sup> We interpret  $x_n$  as being expressed in utils to maintain consistency with existing multi-attribute choice theories. Other theories often make this more explicit by expressing attribute values through a utility function, e.g. letting  $u_n(\mathbf{x})$  denote the value of  $\mathbf{x}$  on attribute  $n$ . We just use  $x_n$  to economize on notation.

<sup>10</sup> We can readily adapt the model to accommodate an attribute, such as price, for which larger values are less desirable by subtracting (instead of adding) the normalized attribute value. For instance, if  $\mathbf{x} = (p_x, q_x)$  and  $\mathbf{y} = (p_y, q_y)$  are defined by their price and a single quality measure, the DM’s normalized valuation of  $\mathbf{x}$  in relation to  $\mathbf{y}$  would simply be  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = \frac{q_x}{q_x + q_y} - \frac{p_x}{p_x + p_y}$ .

<sup>11</sup> Following the literature, here the un-normalized attribute values are implicitly presumed to be separable across dimensions, so that a standard additive preference model,  $V^+(\mathbf{x}) = \sum_n x_n$ , may be regarded as a logical benchmark in the absence of pairwise normalization.  $V^+$  is also typically regarded as a candidate for representing welfare, although this interpretation is not necessary for our analysis.

<sup>12</sup> This does not prevent the model from addressing the context-dependent behaviors described in Section 2. For example, the analysis of the dominance and compromise effects conducted by Noguchi and Stewart (2014) finds that all attributes of all alternatives are typically attended to in two-attribute, three-alternative choice sets.



an additional empirical test of the model.

### 3.1 Binary Choice with Two Attributes

We first consider two-attribute binary choice:

**Lemma 1** *Given  $N = 2$  and  $X = \{\mathbf{x}, \mathbf{y}\}$ ,  $\mathbf{x}$  is preferred to  $\mathbf{y}$  if and only if  $x_1x_2 > y_1y_2$ .*

*Proof of Lemma 1.* From (1),  $\mathbf{x}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}\}$  if and only if  $\frac{x_1}{x_1+y_1} + \frac{x_2}{x_2+y_2} > \frac{y_1}{x_1+y_1} + \frac{y_2}{x_2+y_2}$ . Multiplying through by  $\frac{(x_1+y_1)(x_2+y_2)}{2} > 0$ , then subtracting  $\frac{x_1y_2+y_1x_2}{2}$  from both sides, we see this is equivalent to  $x_1x_2 > y_1y_2$ . ■

Thus, with two attributes, binary-choice preferences under (1) can be equivalently represented by a symmetric Cobb-Douglas preference model,  $V^{\text{CD}}(\mathbf{x}) = x_1x_2$ , which is well-known to generate preferences that are convex and well-behaved. Later in this paper (see Section 7 and Appendix C.5), we will see how some other familiar microeconomic preference representations can relate to parameteric generalizations of the basic PN model.

Much of our subsequent analysis builds on the two-attribute binary-choice problem addressed by Lemma 1. Except where otherwise noted, we will assume that  $\mathbf{x}$  is stronger on the first attribute and  $\mathbf{y}$  is stronger on the second,  $x_1 > y_1$  and  $x_2 < y_2$ , thus ensuring the DM's preference among  $\mathbf{x}$  and  $\mathbf{y}$  is nontrivial. In some cases, it will also be useful to work from the following benchmark of indifference between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice:

**Assumption BCI (Binary-Choice Indifference)** *The DM is indifferent between  $\mathbf{x} = (x_1, x_2)$  and  $\mathbf{y} = (y_1, y_2)$  given  $X = \{\mathbf{x}, \mathbf{y}\}$ .*

Our next result describes how identical improvements to both alternatives on the same dimension will shift preferences towards the alternative that is weaker on that dimension.

**Proposition 1** *Under Assumption BCI, let  $\mathbf{x}' = (x_1, x_2+k)$  and  $\mathbf{y}' = (y_1, y_2+k)$  for some  $k > 0$ . Then  $\mathbf{x}'$  will be preferred to  $\mathbf{y}'$  with  $X = \{\mathbf{x}', \mathbf{y}'\}$ .*

*Proof of Proposition 1.* From Lemma 1,  $\mathbf{x}'$  will be preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}'\}$  if and only if  $x_1(x_2+k) > y_1(y_2+k)$ . Noting  $x_1x_2 = y_1y_2$  under Assumption BCI, this inequality reduces to  $x_1k > y_1k$ , which must hold since  $x_1 > y_1$  and  $k > 0$ . ■

Proposition 1 captures evidence of the relative difference effect (see Section 2). In contrast, most prevailing theories predict that the DM would remain indifferent after both alternatives are improved by the same amount on the same dimension. The exception is Bordalo et al.'s (2013) theory, which allows the preference to shift in both directions.<sup>13</sup>

<sup>13</sup>The direction of this preference shift can depend on many factors, including the extent to which the alternatives are improved (i.e. the magnitude of  $k > 0$ , in our notation). See Appendix D.3 for details.

To aid our interpretation of Proposition 1 (and several later results), we define

$$\Delta(a, b) \equiv \left| \frac{a-b}{a+b} \right|, \quad (2)$$

which provides a metric of the perceptual “distance” or *contrast* between two values, after each value has been normalized in relation to the other.<sup>14</sup> For our two-attribute binary-choice problem, it is then readily verified that  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  if and only if  $\Delta(x_1, y_1) > \Delta(y_2, x_2)$ , in which case there is greater contrast on the dimension where  $\mathbf{x}$  has an advantage than on the dimension where  $\mathbf{y}$  has an advantage.

Given this link between preferences and contrast, Proposition 1 can be understood as arising from a key property of  $\Delta$ , *diminishing sensitivity*, whereby increasing two input values by the same amount decreases the perceived distance between them: in this case,  $\Delta(y_2 + k, x_2 + k) < \Delta(y_2, x_2)$ . The notion that diminishing sensitivity may be important in understanding how individuals perceive value in multi-attribute choice settings was previously highlighted by Bordalo et al. (2013). Along similar lines, we may also regard Proposition 1 as a choice analog of Weber’s (1834) Law of Perception, which describes how increasing the intensities of two stimuli diminishes the perceptibility of their difference — for example, a one-gram difference in the weights of two rocks is more easily detected if the rocks weigh 1 gram and 2 grams than if they weigh 101 grams and 102 grams.

## 4 Adding a Third Alternative to the Choice Set

We now examine how preferences between  $\mathbf{x}$  and  $\mathbf{y}$  may be impacted by adding a third alternative  $\mathbf{z}$ . To aid our understanding, let  $\mathbf{m}^{xy} \equiv (\frac{x_1+y_1}{2}, \frac{x_2+y_2}{2})$  denote the midpoint between  $\mathbf{x}$  and  $\mathbf{y}$ . Noting  $\frac{x_1}{x_2} > \frac{m_1^{xy}}{m_2^{xy}} > \frac{y_1}{y_2}$  (since  $\mathbf{x}$  is better than  $\mathbf{y}$  on the first dimension and worse on the second), we will say that  $\mathbf{z}$  is more *similar* to  $\mathbf{x}$  than to  $\mathbf{y}$  if and only if  $\frac{z_1}{z_2} > \frac{m_1^{xy}}{m_2^{xy}}$ , in which case  $\mathbf{z}$ ’s attribute values are tilted towards  $\mathbf{x}$  and away from  $\mathbf{y}$  in relation to the midpoint.

The next lemma demonstrates the importance of relative similarity in determining the effect of  $\mathbf{z}$  on preferences between  $\mathbf{x}$  and  $\mathbf{y}$ . For ease of exposition, here we consider our benchmark of binary-choice indifference (Assumption BCI), so that a strict preference between  $\mathbf{x}$  and  $\mathbf{y}$  in trinary choice is necessarily caused by the introduction of  $\mathbf{z}$ .

**Lemma 2** *Under Assumption BCI, suppose  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$ . Then, given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,  $\mathbf{x}$  is preferred to  $\mathbf{y}$  if and only if  $\mathbf{x}$  is also preferred to  $\mathbf{z}$ .*

*Proof of Lemma 2.* See Appendix.

<sup>14</sup>This definition of  $\Delta$  — which parallels the standard conceptualization of contrast in the visual perception literature (Carandini and Heeger, 2012) — qualifies as a metric (distance) function because it satisfies: (a)  $\Delta(a, b) \geq 0$  for all  $a, b$ ; (b)  $\Delta(a, b) = 0$  if and only if  $a = b$ ; (c)  $\Delta(a, b) = \Delta(b, a)$ ; and (d)  $\Delta(a, c) \leq \Delta(a, b) + \Delta(b, c)$ . The last property (i.e. the Triangle Inequality) is addressed in Section 5.

Given  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  and with  $\mathbf{x}$  and  $\mathbf{y}$  equally-valued in binary choice, the basic PN model predicts that the magnitude of the perceived value difference will be larger when comparing  $\mathbf{z}$  to  $\mathbf{x}$  than when comparing  $\mathbf{z}$  to  $\mathbf{y}$ :  $|V(\mathbf{x};\{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z};\{\mathbf{x}, \mathbf{z}\})| > |V(\mathbf{y};\{\mathbf{y}, \mathbf{z}\}) - V(\mathbf{z};\{\mathbf{y}, \mathbf{z}\})|$ . In this way, pairwise normalization makes it “easier to compare” more similar alternatives, as proposed by Tversky and Russo (1969) and Natenzon (2018). Consequently, if  $\mathbf{z}$  is inferior to  $\mathbf{x}$  and  $\mathbf{y}$ , it will enhance the perception of  $\mathbf{x}$  more than it enhances the perception of  $\mathbf{y}$ , causing the DM’s preference to shift in favor of  $\mathbf{x}$ .<sup>15</sup> This naturally yields the well-known dominance and compromise effects.<sup>16</sup>

**Proposition 2** *Under Assumption BCI,  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  with  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  in each of the following scenarios:*

- (i)  $\mathbf{x}$  is a compromise between  $\mathbf{y}$  and  $\mathbf{z}$  in that  $z_1 > x_1 > y_1$  and  $y_2 > x_2 > z_2$ , provided  $\mathbf{z}$  is not preferred to  $\mathbf{x}$  and  $\mathbf{y}$ .
- (ii)  $\mathbf{x}$  asymmetrically dominates  $\mathbf{z} \neq \mathbf{x}$  in that  $x_1 \geq z_1 > y_1$  and  $y_2 > x_2 \geq z_2$ .

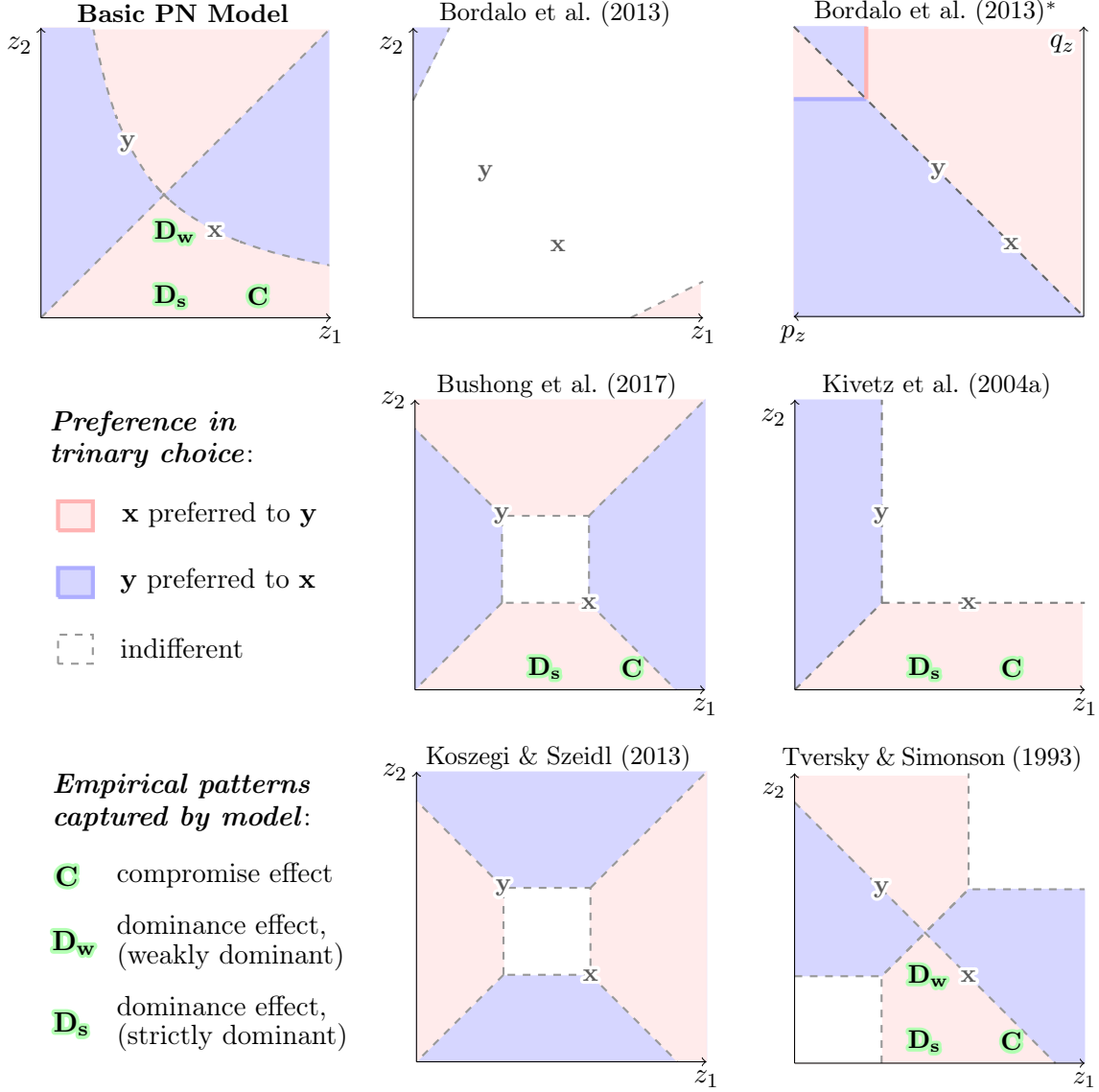
*Proof of Proposition 2.* It is readily apparent that  $\mathbf{x}$  is preferred to  $\mathbf{z}$  in both parts (i) and (ii). Next, observe  $\frac{z_1}{z_2} > \frac{x_1}{x_2} > \frac{m_1^{xy}}{m_2^{xy}}$  given  $z_1 > x_1$  and  $z_2 < x_2$  in part (i), and  $\frac{z_1}{z_2} > \frac{y_1}{x_2} = \frac{m_1^{xy}}{m_2^{xy}}$  given  $z_1 > y_1$  and  $z_2 < x_2$  in part (ii), where  $\frac{y_1}{x_2} = \frac{m_1^{xy}}{m_2^{xy}}$  is verified by cross multiplication with  $x_1 x_2 = y_1 y_2$  (which holds from Lemma 1 with Assumption BCI). The preferences for  $\mathbf{x}$  over  $\mathbf{y}$  with  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  then follow from Lemma 2. ■

The predicted effect of  $\mathbf{z}$  on the preference between  $\mathbf{x}$  and  $\mathbf{y}$  is depicted in the top-left graph in Figure 2. The curved, dashed line on which  $\mathbf{x}$  and  $\mathbf{y}$  reside is the binary-choice indifference curve, where  $\mathbf{z}$  is neither inferior nor superior to  $\mathbf{x}$  and  $\mathbf{y}$ . The dashed line projecting from the origin through the midpoint  $\mathbf{m}^{xy}$  represents the boundary where  $\mathbf{z}$  is equally similar to  $\mathbf{x}$  and to  $\mathbf{y}$ . Therefore, in the lower pink region,  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  and inferior to both, implying  $\mathbf{x}$  is preferred to  $\mathbf{y}$  in trinary choice (Lemma 2) as in the compromise and dominance effects (see the points labeled ‘C,’ ‘D<sub>w</sub>,’ and ‘D<sub>s</sub>’). As shown in the other graphs, prevailing theories do not always predict these effects.

<sup>15</sup> While Lemma 2 does not make reference to the DM’s preference between  $\mathbf{y}$  and  $\mathbf{z}$ , Assumption BCI ensures that the status of  $\mathbf{z}$  as preferred or not preferred to  $\mathbf{x}$  also applies to  $\mathbf{y}$  (and that this holds in both binary and trinary choice). This property is formally established in Appendix C.4. The ease of comparison result as well as a stochastic choice variant of the result are formalized in Appendix C.1.

<sup>16</sup> Note, in both the compromise (part i) and dominance (part ii) effects,  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  (i.e.  $\frac{z_1}{z_2} > \frac{m_1^{xy}}{m_2^{xy}}$ ) and inferior to both. Also note, the requirement for the compromise effect that  $\mathbf{z}$  is not preferred to  $\mathbf{x}$  and  $\mathbf{y}$  fits with empirical demonstrations, as a shift in preference among  $\mathbf{x}$  and  $\mathbf{y}$  will not be observed if  $\mathbf{z}$  is preferred (and hence, chosen) over both. As discussed in Section 4.1, however, the basic PN model’s prediction that the preference shifts *away* from the compromise alternative  $\mathbf{x}$  in relation to  $\mathbf{y}$  when  $\mathbf{z}$  is preferred to both is experimentally testable using a so-called “phantom” choice design.

Figure 2: The Effect of Adding  $\mathbf{z}$  to the Choice Set.



Each graph shows the effect of  $\mathbf{z}$  on the preference between  $\mathbf{x}$  and  $\mathbf{y}$ , as predicted by the indicated model. With one exception (see \*), the graphs were created using  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$  as a simple illustration that ensures binary-choice indifference in all models considered. Additional parametric and functional form restrictions needed to create the graphs are described in Appendix E.

\* While Bordalo et al.'s (2013) model can be analyzed for alternatives defined by two quality attributes (Bordalo et al., 2013, Appendix B), it is primarily analyzed for alternatives defined by their price and a single quality attribute. For this reason, both model variations are considered here, where we use  $\mathbf{x} = (p_x, q_x) = (1, 1)$  and  $\mathbf{y} = (p_y, q_y) = (2, 2)$  to create the 'price-quality' graph.

#### 4.1 The “Strength” of the Compromise and Dominance Effects

Besides simply describing the regions in which  $\mathbf{z}$  will break the DM’s binary-choice indifference in favor of  $\mathbf{x}$  or of  $\mathbf{y}$ , pairwise normalization also provides a measure of where  $\mathbf{z}$  must be located to shift any preference relation. This allows a characterization of how changing  $\mathbf{z}$  can strengthen or weaken its effect on the preference between  $\mathbf{x}$  and  $\mathbf{y}$ .

**Proposition 3** *For  $N = 2$ , suppose the DM prefers  $\mathbf{y}$  to  $\mathbf{x}$  given  $X = \{\mathbf{x}, \mathbf{y}\}$  and is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ . Then  $\mathbf{x}$  will be (strictly) preferred to  $\mathbf{y}$  with  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}$  in each of the following scenarios:*

- (i)  $x_1 > z_1 = z'_1 > y_1$  and  $y_2 > x_2 \geq z_2 > z'_2$ .
- (ii)  $z_1 > x_1 > z'_1 > y_1$  and  $y_2 > x_2 > z_2 = z'_2$ .

*Proof of Proposition 3.* Using  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$  and (1),  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) - V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{(z_2 - z'_2)(y_2 - x_2)(x_2 y_2 - z_2 z'_2)}{(x_2 + z_2)(x_2 + z'_2)(y_2 + z_2)(y_2 + z'_2)}$  in part (i) and  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) - V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{(z_1 - z'_1)(x_1 - y_1)(z_1 z'_1 - x_1 y_1)}{(x_1 + z_1)(x_1 + z'_1)(y_1 + z_1)(y_1 + z'_1)}$  in part (ii). Both expressions must be positive since  $y_2 > x_2 \geq z_2 > z'_2$  and  $z_1 > x_1 > z'_1 > y_1$ , ensuring  $\mathbf{x}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}$ . ■

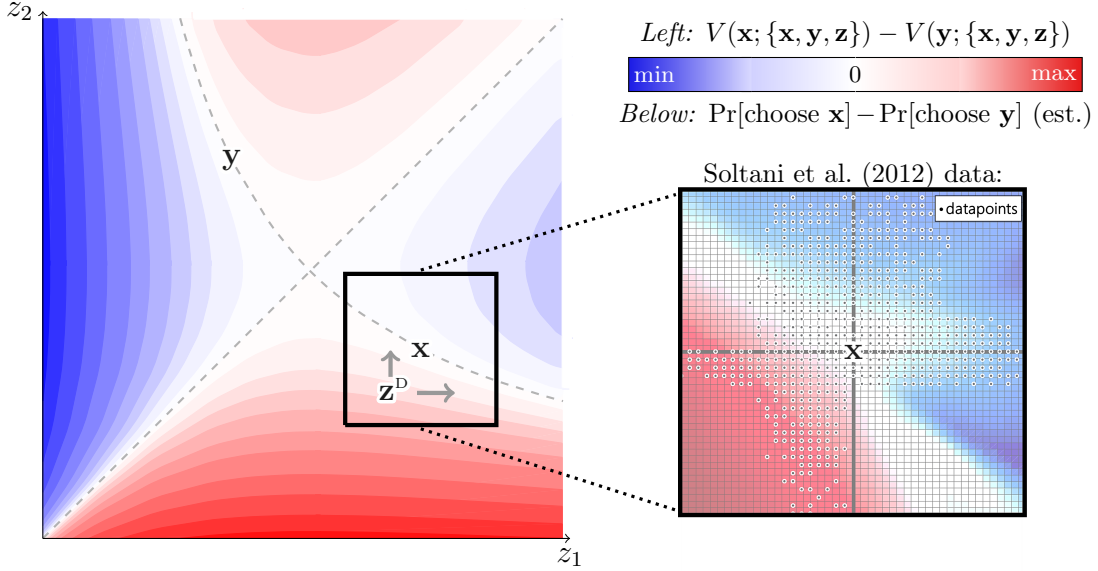
Part (i) of Proposition 3 considers a variation of the dominance effect whereby  $\mathbf{z}$  now causes a shift from a strict binary-choice preference for  $\mathbf{y}$  over  $\mathbf{x}$  to indifference in trinary choice. We then see that an alternate decoy  $\mathbf{z}'$ , which is identical to  $\mathbf{z}$  on the first dimension but worse on the second, instead creates a strict preference for  $\mathbf{x}$ . This prediction fits with evidence of a decoy-range effect (see Section 2), in which the dominance effect becomes more prominent when the decoy becomes worse on the dimension for which it is the weakest alternative in the choice set.

Part (ii) of Proposition 3 analogously considers a variation of the compromise effect whereby  $\mathbf{z}$  causes a shift from a strict binary-choice preference for  $\mathbf{y}$  over  $\mathbf{x}$  to indifference in trinary choice. In turn, an asymmetrically dominated decoy  $\mathbf{z}'$ , which is the same as  $\mathbf{z}$  on the second dimension but is now worse than  $\mathbf{x}$  on the first dimension, will create a strict preference for  $\mathbf{x}$  over  $\mathbf{y}$ . The model therefore predicts that the dominance effect is “stronger” than the compromise effect. While additional tests would be useful, a recent experiment by Noguchi and Stewart (2018) provides preliminary evidence that the dominance effect is indeed stronger than the compromise effect.

To help illustrate how  $\mathbf{z}$ ’s location in attribute space determines the “strength” of its effect on the preference between  $\mathbf{x}$  and  $\mathbf{y}$ , Figure 3 reproduces the graph in Figure 2 depicting the predictions of the basic PN model, except the regions are now shaded based on the magnitude of the difference between the normalized valuations of  $\mathbf{x}$  and  $\mathbf{y}$  in trinary choice. The gray arrows indicate that an asymmetrically-dominated decoy  $\mathbf{z}^D$  enhances the perception of  $\mathbf{x}$  relative to  $\mathbf{y}$ , but this effect weakens as  $z_2^D$  increases — effectively shrinking

the range  $(y_2 - z_2^D)$  of values on this dimension, as in the decoy-range effect — and also as  $z_1^D$  increases to a point where  $\mathbf{x}$  no longer dominates  $\mathbf{z}^D$ , becoming a compromise instead.

Figure 3: The “Strength” of  $\mathbf{z}$ ’s Effect on Preferences Between  $\mathbf{x}$  and  $\mathbf{y}$ .



The graph on left re-illustrates the predicted effect of  $\mathbf{z}$  on the preference between  $\mathbf{x}$  and  $\mathbf{y}$ , as originally shown in Figure 2, except here the regions are shaded based on the magnitude of the normalized value difference between  $\mathbf{x}$  and  $\mathbf{y}$  with  $\mathbf{z}$  included in the choice set. The inset on right provides a visualization of choice data from Soltani et al.’s (2012) experiment, showing  $\mathbf{z}$ ’s effect on the relative likelihood of choosing the more similar alternative  $\mathbf{x}$  relative to the less similar alternative  $\mathbf{y}$  (which were calibrated to reside on the same binary-choice indifference curve). The choice data at each location (pooled over all subjects) is smoothed using a locally-weighted linear regression with smoothing parameter  $1/3$ .

For comparison, the inset in Figure 3 depicts the estimated difference in the choice probabilities of  $\mathbf{x}$  and  $\mathbf{y}$  as a function of the  $\mathbf{z}$ ’s location in attribute space using choice data from Soltani et al. (2012). In this experiment, the locations of  $\mathbf{x}$  and  $\mathbf{y}$  were chosen so that each subject was just indifferent between them (therefore relative valuations, or weights, on each attribute were controlled for). The introduction of  $\mathbf{z}$  necessarily revealed their preference between  $\mathbf{x}$  and  $\mathbf{y}$  because it was only a “phantom” (i.e. it was presented with  $\mathbf{x}$  and  $\mathbf{y}$  but could not actually be chosen). While only suggestive, the observed patterns align with the model’s predictions that increasing  $z_1^D$  and  $z_2^D$  would weaken the effect of  $\mathbf{z}^D$  in shifting preference in favor of  $\mathbf{x}$  relative to  $\mathbf{y}$ . The phantom design also allows us to consider the effect of a superior  $\mathbf{z}$ , which would presumably be chosen over  $\mathbf{x}$  and  $\mathbf{y}$  if it were feasible. In this case, subjects’ preferences appear to shift in favor of  $\mathbf{y}$  instead of  $\mathbf{x}$  (see the blue region above and to the right of  $\mathbf{x}$ ). This matches the prediction of the basic PN model.<sup>17</sup>

<sup>17</sup>Note, while our units (utils) differ from the units of the attributes (which reflected the values of various lotteries) in Soltani et al. (2012), this does not affect the qualitative interpretation of the analysis under

## 5 Binary Choice with Three Attributes

So far, our analysis has only considered choices with alternatives defined on two attribute dimensions. In this section, we consider (binary-choice) preferences among alternatives that vary on three attribute dimensions. Our first such example shows that, with three-attribute choice alternatives, preferences can now be intransitive:

**Example 1** Suppose  $\mathbf{x} = (a, b, c)$ ,  $\mathbf{x}' = (b, c, a)$ , and  $\mathbf{x}'' = (c, a, b)$  with  $a > b > c$ . Then:

$$\begin{aligned} V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) - V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) &= V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) - V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\}) \\ &= V(\mathbf{x}''; \{\mathbf{x}, \mathbf{x}''\}) - V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}''\}) = \Delta(a, b) + \Delta(b, c) - \Delta(a, c) = \frac{(a-b)(b-c)(a-c)}{(a+b)(b+c)(a+c)} > 0. \end{aligned}$$

Thus, in binary choices,  $\mathbf{x}$  is preferred to  $\mathbf{x}'$ ,  $\mathbf{x}'$  is preferred to  $\mathbf{x}''$ , and  $\mathbf{x}''$  is preferred to  $\mathbf{x}$ .

In Example 1,  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  satisfy a ‘cyclical majority-dominance’ property whereby  $\mathbf{x}$  is better than  $\mathbf{x}'$  on two of three attributes,  $\mathbf{x}'$  is better than  $\mathbf{x}''$  on two of three attributes, and  $\mathbf{x}''$  is better than  $\mathbf{x}$  on two of three attributes. In turn, the DM exhibits a majority-rule preference cycle, as each alternative is preferred to that for which it is better on two out of three attributes. This particular preference cycle arises directly from the fact that, as a metric of perceptual distance, the contrast function satisfies the triangle inequality (see footnote 14). That is, if  $a > b > c > 0$ , then

$$\Delta(a, b) + \Delta(b, c) > \Delta(a, c).$$

This relation implies that, for any two alternatives among  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  in Example 1, the total contrast on the two dimensions for which the majority-dominant alternative has an advantage will be greater than the contrast on the dimension for which the minority-dominant alternative has an advantage.

In general, with three potential alternatives that satisfy the cyclical majority-dominance property, PN does **not** necessarily imply that all binary-choice preferences are intransitive. Instead, it predicts that if they are intransitive, they will only be intransitive in one direction. In principle, intransitivity could arise in one of two ways: a majority-rule preference cycle (as in Example 1); or an opposite ‘minority-rule’ cycle. The next result clarifies that only majority-rule preference cycles can arise under PN, which matches evidence discussed in Section 2:

**Proposition 4** Given  $N = 3$ , suppose  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  satisfy  $x_1 > x'_1 > x''_1$ ,  $x''_2 > x_2 > x'_2$ , and  $x'_3 > x''_3 > x_3$ . Then, if binary-choice preferences among  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  are intransitive, it must be the case that  $\mathbf{x}$  is preferred to  $\mathbf{x}'$ ,  $\mathbf{x}'$  is preferred to  $\mathbf{x}''$ , and  $\mathbf{x}''$  is preferred to  $\mathbf{x}$ .

---

the natural assumption that these units are positively related.

*Proof of Proposition 4.* See Appendix.

Our next result considers the effect of splitting an attribute into two sub-attributes, effectively re-framing a choice between two-attribute alternatives,  $\mathbf{x}$  and  $\mathbf{y}$ , as a choice between three-attribute alternatives,  $\mathbf{x}'$  and  $\mathbf{y}'$ :

**Proposition 5** *Under Assumption BCI, let  $\mathbf{x}' = (x_{1a}, x_{1b}, x_2)$  and  $\mathbf{y}' = (y_{1a}, y_{1b}, y_2)$  with  $x_{1a} + x_{1b} = x_1$ ,  $y_{1a} + y_{1b} = y_1$ ,  $x_{1a} \geq y_{1a}$ , and  $x_{1b} \geq y_{1b}$ . Then  $\mathbf{x}'$  is preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}'\}$ .*

*Proof of Proposition 5.*  $\mathbf{x}'$  will be preferred to  $\mathbf{y}'$  if and only if  $\Delta(x_{1a}, y_{1a}) + \Delta(x_{1b}, y_{1b}) > \Delta(y_2, x_2)$ . Under Assumption BCI,  $\Delta(y_2, x_2) = \Delta(x_1, y_1) = \Delta(x_{1a} + x_{1b}, y_{1a} + y_{1b})$ , which implies the previous condition is equivalent to  $\Delta(x_{1a}, y_{1a}) + \Delta(x_{1b}, y_{1b}) > \Delta(x_{1a} + x_{1b}, y_{1a} + y_{1b})$ , which itself is equivalent to  $\frac{(x_{1a} - y_{1a})(x_{1b} + y_{1b})^2 + (x_{1b} - y_{1b})(x_{1a} + y_{1a})^2}{(x_{1a} + y_{1a})(x_{1b} + y_{1b})(x_{1a} + y_{1a} + x_{1b} + y_{1b})} > 0$  and must hold since both terms in the numerator are non-negative and at least one is strictly positive given  $x_{1a} \geq y_{1a}$  and  $x_{1b} \geq y_{1b}$  (with at most one inequality binding). ■

Consistent with evidence of the splitting bias (see Section 2), attribute-splitting shifts preferences toward the alternative that is stronger on the split attribute, in this case  $\mathbf{x}$ , provided its advantage is maintained on each sub-attribute. The reason  $\mathbf{x}$ 's advantage over  $\mathbf{y}$  on attribute 1 is perceived to be larger when spread over two sub-attributes naturally follows from the fact that the contrast function  $\Delta$  satisfies the triangle inequality.<sup>18</sup>

Next, we examine the effect of attribute alignability on binary-choice preferences. Here, an attribute is considered ‘alignable’ if the corresponding attribute values are strictly positive for both alternatives. To isolate the effect of alignability, we will work from Assumption BCI, while presuming that both attributes are alignable. We then consider preferences among two modified alternatives with only one alignable attribute:

**Proposition 6** *Under Assumption BCI with  $\min\{x_1, x_2, y_1, y_2\} > 0$ , let  $\mathbf{x}' = (x_1, x_2, 0)$  and  $\mathbf{y}' = (y_1, 0, y_2)$ . Then  $\mathbf{x}'$  is preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}'\}$ .*

*Proof of Proposition 6.*  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) - V(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\}) = \frac{x_1 - y_1}{x_1 + y_1} + \frac{x_2}{x_2} - \frac{y_2}{y_2} = \frac{x_1 - y_1}{x_1 + y_1} > 0$ . ■

The preference for  $\mathbf{x}'$  over  $\mathbf{y}'$  described by Proposition 6 (along with indifference between  $\mathbf{x}$  and  $\mathbf{y}$ ) indicates that the advantage  $y_2 > x_2$  is weighted more heavily if  $y_2$  and  $x_2$  exist on the same attribute dimension than if they exist on separate (i.e. non-alignable) dimensions. This matches evidence of the alignability effect described in Section 2.

<sup>18</sup>This effect is also amplified by diminishing sensitivity in  $\Delta$ . To illustrate, suppose  $x_1 = 6$ ,  $y_1 = 4$ ,  $x_{1a} = x_{1b} = 3$ , and  $y_{1a} = y_{1b} = 2$ . The triangle inequality implies  $\Delta(6, 4) < \Delta(6, 5) + \Delta(5, 4)$ , while diminishing sensitivity implies  $\Delta(6, 5) < \Delta(5, 4) < \Delta(3, 2)$ . Thus,  $\Delta(6, 4) < \Delta(3, 2) + \Delta(3, 2)$ , which means the total contrast between  $x_{1a}$  and  $y_{1a}$  and between  $x_{1b}$  and  $y_{1b}$  exceeds the contrast between  $x_1$  and  $y_1$ .



## 6 Allocation Problems

We now consider a choice between different allocations of an asset (or resource), with total value  $A > 0$ , across  $N$  dimensions. A given allocation  $\mathbf{x}$  of the asset then satisfies  $\sum_{n \leq N} x_n = A$ . For simplicity, this setup implicitly presumes that allocations generate the same rate of return on all dimensions. Unequal returns are considered in Appendix C.2.

While stylized, the formulation described above provides a simple baseline that can be related to a variety of allocation problems. For example,  $A$  could represent a budget that is spent on consumption bundles defined over  $N$  goods or an investor's recurring contribution to a savings plan that includes  $N$  different funds.

**Proposition 7** *Given  $N > 1$  and  $A > 0$ , suppose  $x_n = \frac{A}{N}$  for all  $n \leq N$ . Then, for any  $\mathbf{x}' \neq \mathbf{x}$  with  $\sum_{n \leq N} x'_n = A$ ,  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ .*

*Proof of Proposition 7.* See Appendix.

From Proposition 7, a balanced allocation that allocates an equal  $\frac{1}{N}$  share of the asset to each dimension will be strictly preferred to any other possible allocation of the asset. This result aligns with evidence of a diversification bias, such as Benartzi and Thaler's (2001) finding that investors often follow a " $\frac{1}{N}$  heuristic" by selecting a balanced savings plan that allocates contributions equally across the  $N$  available funds. Note, since we abstract from the possibility of uncertain returns, this preference for a balanced allocation cannot be rationalized as a variance-reduction strategy and thus represents a "bias" in relation to a standard additive preference model ( $V^+(\mathbf{x}) = \sum_n x_n$ ), which would predict indifference between any two allocations of the same asset. Furthermore, Proposition 7 still applies even if allocations yield higher returns on some dimensions than others (see Appendix C.2), in which case the interpretation as a "bias" may be more self-evident.

Considering this preference for a balanced allocation, it is natural to suspect that a DM would generally favor allocations for which all dimensions receive a positive share of the asset. To explore this idea, suppose two firms previously offered identical products defined on  $N - 1 > 0$  dimensions, each of which may be thought of as representing a distinct product feature. However, both firms have since invested  $q > 0$  in research and development to improve their products. One firm improved the quality (i.e. unnormalized attribute value) of an existing product feature, from  $x_{n'} > 0$  to  $x_{n'} + q$ , on dimension  $n' \leq N - 1$ . The other firm innovated a  $N$ th product feature, attaining a quality level of  $x_N = q$  on this new dimension. As our next result shows, the product with the new feature will now be preferred to the product with the improvement to an existing feature:

**Proposition 8** *Given  $N > 1$  and  $q > 0$ , suppose  $x_N = q$ ,  $x'_N = 0$ ,  $x'_{n'} = x_{n'} + q$  for some  $n' < N$ , and  $x'_n = x_n > 0$  for all other  $n < N$ . Then  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ .*

*Proof of Proposition 8.*  $V(\mathbf{x}; X) - V(\mathbf{x}'; X) = \Delta(q, 0) - \Delta(x_{n'} + q, x_{n'}) = \frac{2x_{n'}}{q+2x_{n'}} > 0$ . ■

This prediction that a new product feature will be valued more than an otherwise-equivalent improvement to an existing product feature fits with evidence of the feature bias (see Section 2). Here, the feature bias can be understood as a consequence of diminishing sensitivity in  $\Delta$ . Since the mean attribute value between the two products is higher for the existing feature than for the new feature (i.e.  $x_{n'} + \frac{q}{2} > \frac{q}{2}$ ), the value difference on the new dimension  $N$  will, as a result of diminishing sensitivity, be perceived as greater than the equal-sized value difference on the existing dimension  $n'$ .

Proposition 8 applies even if allocations yield higher returns on some dimensions than others (see Appendix C.2). This is especially noteworthy if the return to investing  $q$  (in terms of the increase in the corresponding attribute value) is lower on the new dimension. In this case, the product with the new feature would be preferred despite its lower overall quality. Thus, in product-level investment decisions, firms would naturally have an incentive to allocate research and development resources towards innovating new features, even if they add little actual value to the product. In this way, pairwise normalization offers a potential explanation for the phenomenon known as “feature bloat” or “feature creep” — i.e. the proliferation of products with an excessive number of features (e.g. Thompson and Norton, 2011) — as well as the related observation that developing “irrelevant” new product features can foster a sustained competitive advantage (Carpenter et al., 1994).

## 7 A One-Parameter Generalization

This section considers a generalization of our model based on a common one-parameter formulation of the normalization computation in neuroscience. The exercise will allow us to assess the extent to which the predictions of the basic model are maintained under this generalization, while also revealing some new predictions.

In this formulation, the normalized value of a single input  $a$  in relation to  $b$  is now  $\frac{a}{\sigma+a+b}$  with  $\sigma \geq 0$ . As with the  $\frac{a}{a+b}$  model, we adapt the  $\frac{a}{\sigma+a+b}$  model to multi-attribute choice through our notion of pairwise, attribute-level comparisons as

$$V^*(\mathbf{x}; X) = \sum_{n \leq N} \sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{x_n}{\sigma + x_n + y_n}. \quad (3)$$

A more detailed discussion of how  $\sigma$  affects the perception of attributes in the model, and its foundation in the neuroscience literature, is provided in Appendix C.5.

### 7.1 Binary-Choice Preferences

Binary-choice preferences with two attributes can now be represented as a composite of preferences under the (symmetric) Cobb-Douglas and additive preference models.

**Lemma 3** Given  $V^{\text{CD}}(\mathbf{x}') = x_1''x_2''$  and  $V^+(\mathbf{x}'') = x_1'' + x_2''$  with  $N = 2$  and  $\mathbf{x}'' \in \{\mathbf{x}, \mathbf{x}'\}$ :

(i) If  $V^{\text{CD}}(\mathbf{x}) \geq V^{\text{CD}}(\mathbf{x}')$ ,  $V^+(\mathbf{x}) \geq V^+(\mathbf{x}')$ , and  $\sigma \geq 0$  with at least two of these inequalities non-binding, then  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ .

(ii) If  $V^{\text{CD}}(\mathbf{x}) > V^{\text{CD}}(\mathbf{x}')$  and  $V^+(\mathbf{x}') > V^+(\mathbf{x})$ , then  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$  if and only if  $\sigma < \frac{2(V^{\text{CD}}(\mathbf{x}) - V^{\text{CD}}(\mathbf{x}'))}{V^+(\mathbf{x}') - V^+(\mathbf{x})}$ .

*Proof of Lemma 3.* See Appendix.

Thus, if the Cobb-Douglas ( $V^{\text{CD}}$ ) and additive ( $V^+$ ) preference models agree in their rankings among two alternatives, the DM's preference will align with this ranking. If there is disagreement, the DM's preference will coincide with the additive model if and only if  $\sigma$  is sufficiently large. Thus, a larger  $\sigma$  effectively implies a larger weight on additive relative to Cobb-Douglas preferences in determining the preference under (3).

The role of  $\sigma$  in arbitrating binary-choice preferences is depicted in Figure 4. Compared to the basic PN model (equivalently represented by Cobb-Douglas preferences, top left), the model with  $\sigma > 0$  predicts flatter indifference curves (top right).<sup>19</sup> In the large- $\sigma$  limit, binary choice preferences converge to additive preferences, which also describes binary-choice preferences in most prevailing multi-attribute choice theories (bottom right).

## 7.2 Robustness of Key Behavioral Predictions

Next, we see that many key predictions of the basic PN model are maintained with  $\sigma > 0$ :

**Proposition 9** For all  $\sigma \geq 0$ , the following results still hold under (3):

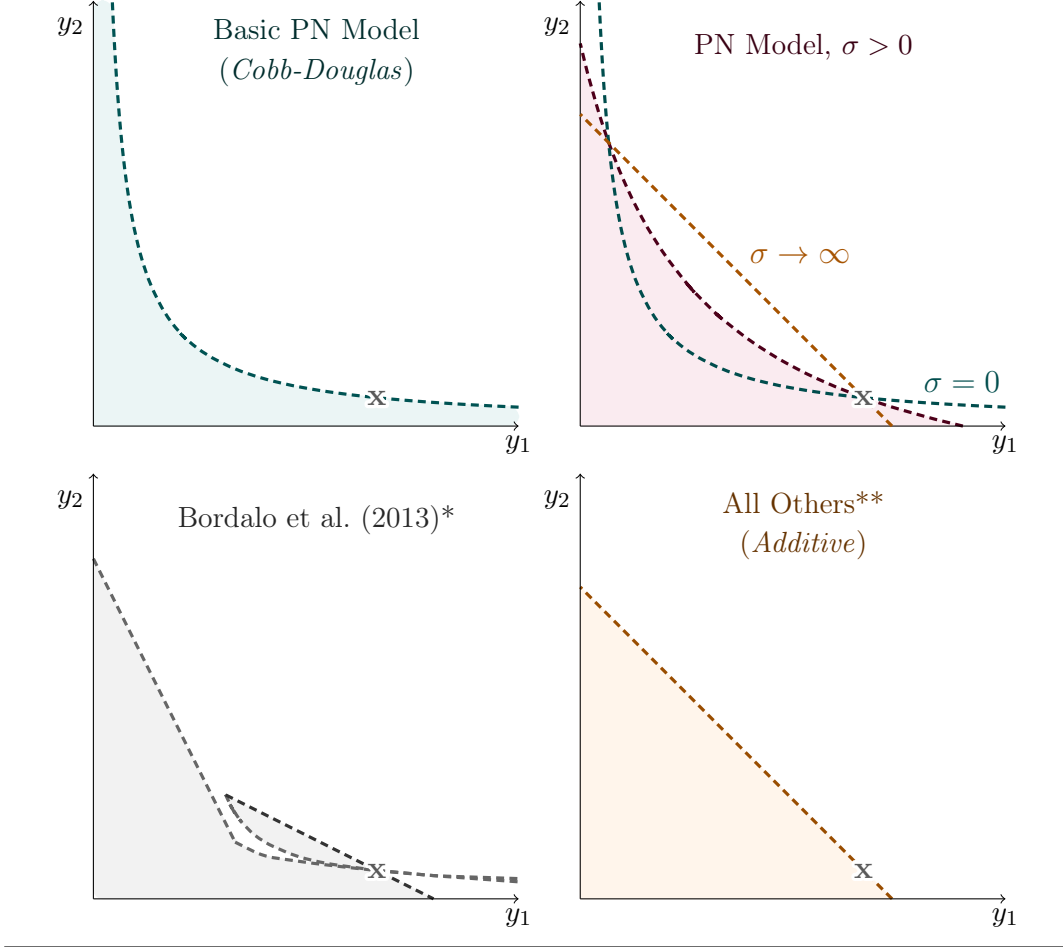
- (i) the relative difference effect (Proposition 1);
- (ii) majority-rule preference cycles (Proposition 4);
- (iii) the splitting bias (Proposition 5);
- (iv) the alignability effect (Proposition 6);
- (v) the diversification bias (Proposition 7);
- (vi) the feature bias (Proposition 8).

*Proof of Proposition 9.* See Appendix.

To clarify the one-parameter model's predictions regarding the compromise and dominance effects, which are not addressed in Proposition 9, we will now examine the effect of adding a third alternative  $\mathbf{z}$  on preferences between  $\mathbf{x}$  and  $\mathbf{y}$  under (3), and how this effect may depend on the magnitude of  $\sigma > 0$ . To do this, we will again work from a benchmark

<sup>19</sup>In addition to Cobb-Douglas and additive preferences, pairwise normalization can also be related to constant elasticity of substitution (CES) preferences and rank-based lexicographic preferences, both of which are nested as special cases of a two-parameter version of our model. See Appendix C.5.

Figure 4: Binary-Choice Preferences.



The shaded region(s) of each graph indicate the region(s) where  $\mathbf{x}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}\}$  with  $N = 2$ , as predicted by the indicated model(s) with  $\mathbf{x} = (1, .1)$ .

\* Bordalo et al.'s (2013) model can generate many geometric configurations of the binary-choice preference regions. For example, the shapes of the regions can vary with a 'salience distortion' parameter  $\delta \in (0, 1)$  (this illustration uses  $\delta = .5$ ) as well as the choice of  $\mathbf{x}$  (e.g. using a different  $\mathbf{x}$  on the boundary between the shaded and unshaded regions can yield different preference regions than those shown here). For an exact description of the version of Bordalo et al.'s (2013) model used to generate this graph, see Appendix E.

\*\* Includes the three remaining comparable models from Table 1: Tversky and Simonson (1993), Kivetz et al. (2004a), and Bushong et al. (2017), as well as Koszegi and Szeidl (2013). Again, see Appendix D and Appendix E for additional details on these models.

of indifference between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice. An added complication, however, is that allowing  $\sigma$  to vary may undo binary-choice indifference in light of Lemma 3. For this reason, we will adopt a stronger version of Assumption BCI, which ensures binary-choice indifference is preserved even as  $\sigma$  varies:

**Assumption BCI\*** *The DM is indifferent between  $\mathbf{x} = (x_1, x_2)$  and  $\mathbf{y} = (y_1, y_2)$  given  $X = \{\mathbf{x}, \mathbf{y}\}$  for all  $\sigma \geq 0$  under (3) (equivalently,  $x_1 = y_2$  and  $y_1 = x_2$ ).*

**Lemma 4** *Under Assumption BCI\* and given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ , if  $\mathbf{x}$  is preferred to  $\mathbf{y}$  with  $\sigma = 0$ , then  $\mathbf{x}$  will still be preferred to  $\mathbf{y}$  with  $\sigma > 0$  unless  $V^{\text{CD}}(\mathbf{z}) < V^{\text{CD}}(\mathbf{x}) < V^{\text{CD}}(\mathbf{z}')$ , where  $\mathbf{z}' \equiv (z_1 + \sigma, z_2 + \sigma)$ .*

*Proof of Lemma 4.* See Appendix.

Lemma 4 states that the trinary-choice preference between  $\mathbf{x}$  and  $\mathbf{y}$  predicted by the basic PN model may be reversed with  $\sigma > 0$ , but only in the event that  $\mathbf{z}$  is “modestly inferior” to  $\mathbf{x}$  in the sense that  $\mathbf{x}$  is preferred to  $\mathbf{z}$  according to the Cobb-Douglas preference model, but would not be preferred to some  $\mathbf{z}'$  featuring a magnitude- $\sigma$  improvement to  $\mathbf{z}$  on each dimension (see Figure 5). It follows that the compromise and dominance effects still hold with  $\sigma > 0$ , unless  $\mathbf{z}$  is only modestly inferior, in which case the opposite trinary-choice preference — favoring  $\mathbf{y}$  over  $\mathbf{x}$  — is predicted:

**Proposition 10** *Under Assumption BCI\* and with  $\mathbf{z}'$  defined as in Lemma 4, suppose  $\mathbf{z}$  satisfies the conditions for either the compromise or dominance effect in Proposition 2, implying  $\mathbf{x}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  with  $\sigma = 0$ . With  $\sigma > 0$ ,  $\mathbf{x}$  is still preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  if  $V^{\text{CD}}(\mathbf{x}) > V^{\text{CD}}(\mathbf{z}')$ , but  $\mathbf{y}$  is preferred to  $\mathbf{x}$  if  $V^{\text{CD}}(\mathbf{x}) < V^{\text{CD}}(\mathbf{z}')$ .*

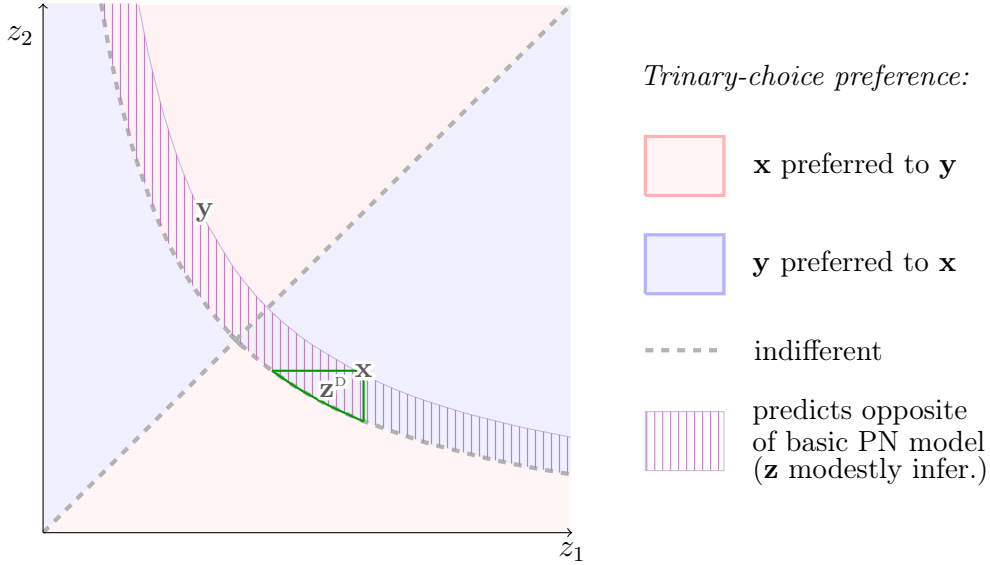
*Proof of Proposition 10.* See Appendix.

To aid our understanding of this result, consider a modestly inferior decoy  $\mathbf{z}^{\text{D}}$ , that is asymmetrically dominated by  $\mathbf{x}$ . Since it is only modestly inferior,  $\mathbf{z}^{\text{D}}$  will not just be more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  (as described in Section 4), it will also be sufficiently similar to  $\mathbf{x}$  in an absolute sense — i.e. its proximity in attribute space. In Figure 5, for example, this decoy must reside in the region bounded by the green lines in close proximity of  $\mathbf{x}$ .

Proposition 10’s implication that  $\mathbf{z}^{\text{D}}$  would create a preference for the *dissimilar* alternative  $\mathbf{y}$  instead of  $\mathbf{x}$  — in apparent opposition of the dominance effect — fits with Tversky’s (1972) “similarity hypothesis.”<sup>20</sup> Though the dominance effect is still widely-accepted as

<sup>20</sup> Unlike the dominance effect, a reversal of the compromise effect with  $\sigma > 0$  and a modestly inferior  $\mathbf{z}$  is not necessarily observable. This is because “modest inferiority” is based on Cobb-Douglas preferences, which do not necessarily align with preferences under (3). Consequently, a modestly inferior  $\mathbf{z}$  may actually be preferred to  $\mathbf{x}$  and  $\mathbf{y}$  in trinary choice with  $\sigma > 0$ , in which case a preference for  $\mathbf{y}$  over  $\mathbf{x}$  may be concealed as  $\mathbf{z}$  would be selected over both.

Figure 5: The Effect of a Third Alternative  $\mathbf{z}$  when  $\sigma > 0$ .



This graph shows the effect of adding  $\mathbf{z}$  to the choice set on the DM's preference between  $\mathbf{x}$  and  $\mathbf{y}$ , as predicted by the one-parameter PN model. The green lines denote the boundaries of the 'similarity' subregion, as described in the text.

a robust empirical regularity (e.g. Huber et al., 2014), some support for the similarity hypothesis comes from Frederick et al.'s (2014) recent unincentivized choice experiment. In this study, the share of subjects who preferred the dominant (and more similar) alternative increased by more than one percent in just one of the eleven product classes considered — and actually decreased in most classes — when an asymmetrically-dominated decoy was added to the choice set.

While research investigating the precise boundaries of the dominance effect is ongoing, the one-parameter model allows for the possibility that the dominance effect might not arise in all situations. Our next result offers more specific guidance on this point, showing that the dominance effect must hold with sufficiently high-value goods:

**Corollary 1** *Under Assumption BCI\*, with  $\sigma > 0$  and  $\mathbf{z}$  asymmetrically dominated by  $\mathbf{x}$ , suppose the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ . Also let  $\mathbf{w}' = (\gamma \cdot w_1, \gamma \cdot w_2)$  for each  $\mathbf{w} \in \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ . Then,  $\mathbf{x}'$  is preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}', \mathbf{z}'\}$  if  $\gamma > 1$ , while  $\mathbf{y}'$  is preferred to  $\mathbf{x}'$  if  $\gamma < 1$ .*

*Proof of Corollary 1.* See Appendix.

Corollary 1 first considers a benchmark in which the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in both binary and trinary choice — here,  $\mathbf{z}$  neither helps nor hurts the perception of  $\mathbf{x}$  relative to  $\mathbf{y}$ . In turn,  $\mathbf{x}'$ ,  $\mathbf{y}'$ , and  $\mathbf{z}'$  are defined as analogs to  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{z}$ , except with

their attribute values scaled by a constant  $\gamma > 0$ . As seen,  $\mathbf{x}'$  will then be preferred to  $\mathbf{y}'$  in trinary choice with  $\mathbf{z}'$  — consistent with the dominance effect — if (and only if)  $\gamma > 1$ , in which case  $\mathbf{x}'$ ,  $\mathbf{y}'$ , and  $\mathbf{z}'$  represent higher-value alternatives than  $\mathbf{x}$ ,  $\mathbf{y}$ , and  $\mathbf{z}$ .

Corollary 1’s implication that the dominance effect will be more prominent for higher-value goods has support in the empirical literature. For example, in Frederick et al.’s (2014) study mentioned above, the lone product class in which the decoy created a non-negligible shift in subjects’ preferences towards the dominant option was also the highest-value product class considered in their study; in Huber et al.’s (1982) original study documenting the dominance effect, the decoy shifted subjects’ preferences in favor of the dominant alternative in all six product classes considered, but the largest effects were likewise observed in the two highest-value product classes.<sup>21</sup> In fact, a study by Malkoc et al. (2013) directly manipulated alternatives’ desirability within each product class. A robust dominance effect was observed with more desirable alternatives, but not with less desirable alternatives. As the authors conclude, their results “establish (un)desirability as an important boundary condition” for the dominance effect, as Corollary 1 would suggest.

## 8 “Comparisons” in Multi-Attribute Choice Theories

This paper presented a theory of multi-attribute choice based on a notion of pairwise attribute-level comparisons, implemented by (divisive) normalization — a well-documented and normatively-grounded form of relative value encoding used in neural processing. As we have shown, pairwise normalization can explain a wide range of context-dependent behaviors, including several that are not well-addressed in the theoretical literature (again, see Table 1).

Like our model, other multi-attribute choice theories typically suggest that an alternative’s attributes are “compared” (or otherwise valued in relation) to the corresponding attributes of other alternatives. While formal representations of attribute-level comparisons vary from model to model, the use of the normalization computation for this general purpose is not unique to our theory. For instance, Bordalo et al. (2013)’s proposed form of their “salience function” (eq. 4, p. 809) is identical to our contrast function  $\Delta(a, b) = \left| \frac{a-b}{a+b} \right|$ . However, their implementation differs. While we use  $\Delta$  to express the perceived, decision-relevant value difference between two attribute values, Bordalo et al.’s salience function is used to compare an attribute value to the average over the choice set (along that dimension), and as an intermediate step in computing the perceived value of that attribute.

Tversky and Simonson’s (1993) model also uses normalization to express the total

---

<sup>21</sup> The highest-value product class considered by Frederick et al. (2014) was apartments, while the others were bottled water, fruit, hotel rooms, jelly beans, kool-aid, mints, movies, and popcorn; the two highest-value product classes considered by Huber et al. (1982) were cars and televisions, while the others were beer, photographic film, lotteries (with expected payoffs on the order of \$20), and restaurant meals.

“advantage” of  $\mathbf{x}$  over  $\mathbf{y}$ ,  $A(\mathbf{x}, \mathbf{y}) = \sum_n \max\{x_n - y_n, 0\}$  relative to its “disadvantage,”  $D(\mathbf{x}, \mathbf{y}) = A(\mathbf{y}, \mathbf{x})$  in its simplest form, as  $\frac{A(\mathbf{x}, \mathbf{y})}{A(\mathbf{x}, \mathbf{y}) + D(\mathbf{x}, \mathbf{y})}$  (eq. 8, p. 1185). Although this computation is applied only after attribute information is aggregated across dimensions (unlike our use of normalization), Tversky and Simonson conceptualize the advantage and disadvantage functions as arising from pairwise comparisons of attribute values.<sup>22</sup>

Table 2 describes these shared features as well as other representations of attribute-level comparisons used in the prevailing multi-attribute choice theories listed in Table 1. As seen, attribute-level comparisons are either implemented through normalization, through the use of attribute-weights, or through subtraction (with possible additional transformations). In these comparisons, attribute values are either compared to each other (in pairs) or to a summary statistic, such as the average, minimum, or range of attribute values on that dimension.

Table 2. “Comparisons” in Multi-Attribute Choice Models

	<i>Attribute-level “comparisons”...</i>		Inter-attribute “comparisons” of attribute-level outputs?
	Computation used in each comparison	What is each attribute value compared to?	
<b>Pairwise Normalization</b>	<b><i>normalization</i></b>	<b><i>other attribute values, in pairs</i></b>	no
Bordalo et al. (2013)	<b><i>normalization*</i></b>	average of attribute values	yes, outputs are ranked
Bushong et al. (2017)	weight by decr. function of...	range of attribute values (max - min)	no
Kivetz et al. (2004a)	subtraction**	minimum of attribute values	no
Koszegi and Szeidl (2013)	weight by incr. function of...	range of attribute values (max - min)	no
Tversky and Simonson (1993)	subtraction	<b><i>other attribute values, in pairs</i></b>	yes, through <b><i>normalization*</i></b>

\* See text for relevant caveats and Appendix D for technical details.

\*\* Additional transformations of the difference between two attribute values may be applied.

Regarding the objects of attribute-level comparisons, there are certainly advantages and disadvantages to each approach. Our pairwise formulation, in which attribute values are compared directly to other attribute values, is largely motivated by eye-tracking evidence (discussed in Section 1). It is also worth considering the computational demands

<sup>22</sup> Also see Marley (1991) and Tserenjigmid (2015) for axiomatizations of pairwise comparisons.



of each approach. For binary-choice problems, our approach is certainly simple, as it only requires a single computation to express the perceived value of an attribute, i.e.  $\frac{x_n}{x_n + y_n}$ , while comparisons to summary statistics would require at least two distinct computations — the computation of the summary statistic itself and the computation used to implement the comparison between the attribute value and that summary statistic. With many alternatives, however, the use of a summary statistic could certainly simplify the problem. This observation reinforces the sentiment (expressed in Section 3) that in choice environments where there are too many alternatives to realistically carry out every possible pairwise comparison, the model would have different predictions. Therefore the reported results should be interpreted as applying to settings when all attributes and alternatives are considered.<sup>23</sup>

Models that entail attribute-level comparisons to a summary statistic are naturally equipped to address evidence that preferences can be sensitive to that particular statistic. For instance, behavior dependent on the average attribute value is evident from empirical evidence of the relative difference effect (see Section 2). Experimental research has also revealed range-dependence, whereby a fixed difference between two attribute values is weighted less when the range of attribute values on that dimension is wider (Mellers and Cooke, 1994; Yeung and Soman, 2005). Although pairwise normalization does not embed attribute-level comparisons to the average or range of attribute values in the choice set, it nonetheless captures both average- and range-dependence in choice, as Proposition 1 established average-dependence (in the form the relative difference effect) while range-dependence is demonstrated in Appendix C.3.

As mentioned in Section 4, pairwise normalization also implies that more similar alternatives will be “easier to compare” than less similar alternatives.<sup>24</sup> This idea is also prominent in Natenzon’s (2018) Bayesian probit model, in which an imperfectly-informed (yet rational) decision-maker can exhibit the compromise and dominance effects due to the relative ease of comparing an inferior third alternative to the existing alternative to which it is more similar. In Natenzon’s model, the ease of comparison concept is operationalized as an assumption about value correlations among alternatives that may be encountered in one’s environment. Pairwise normalization not only provides a potential foundation for

---

<sup>23</sup> Of course, with arbitrarily many alternatives and/or attributes, it may also be impractical to compute summary statistics across all alternatives on every dimension, not to mention carrying out the additional intra-attribute comparisons embedded in some models (see Table 2). With that said, Koszegi and Szeidl’s (2013) focusing theory may be regarded as providing a reduced-form description of how decision-makers allocate attention across attributes when there are many possible attributes to consider (as opposed to providing an explanation for the context effects addressed by other theories).

<sup>24</sup> Formally, if  $\mathbf{x}$  and  $\mathbf{y}$  are equally-valued in binary choice, and  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  (in that its attributes are “tilted” towards  $\mathbf{x}$  and away from  $\mathbf{y}$  relative to their midpoint), then the (absolute) perceived value difference will be larger — suggesting a “more conspicuous” preference — between  $\mathbf{x}$  and  $\mathbf{z}$  than between  $\mathbf{y}$  and  $\mathbf{z}$ . See Appendix C.1 for a formal demonstration of this relationship as well as an observable, stochastic choice version of the result.

Natenzon's assumption, it also indicates that the relative ease of comparing more similar alternatives does not need to reflect an inherent feature of the alternatives themselves. Instead, our model suggests that the manner in which our brains encode sensory information effectively makes it easier to compare more similar alternatives.

## References

- [1] Arieli, Amos, Ben-Ami, Yaniv, and Rubinstein, Ariel, “Tracking Decision Makers Under Uncertainty,” *American Economic Journal: Microeconomics*, 3 (2011), 68–76.
- [2] Azar, Ofer, “The Effect of Relative Thinking on Firm Strategy and Market Outcomes: A Location Differentiation Model with Endogenous Transportation Costs,” *Journal of Economic Psychology*, 29 (2008), 684–697.
- [3] Bardolet, David, Fox, Craig, and Lovallo, Dan, “Naive Diversification and Partition Dependence in Capital Allocation Decisions,” unpublished manuscript (2007).
- [4] Bateman, Ian, Munro, Alistair, Rhodes, Bruce, Starmer, Chris, and Sugden, Robert, “Does Part–Whole Bias Exist? an Experimental Investigation,” *The Economic Journal*, 107 (1997), 322–332.
- [5] Benartzi, Shlomo, and Thaler, Richard, “Naive Diversification Strategies In Defined Contribution Saving Plans,” *American Economic Review*, 91 (2001), 79–98.
- [6] Bernheim, Douglas, “On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal,” *American Economic Journal: Microeconomics*, 1 (2009), 1–41.
- [7] Birnbaum, Michael, and Bahra, Jeffrey “Gain-Loss Separability and Coalescing in Risky Decision Making,” *Management Science*, 53 (2007), 1016–1028.
- [8] Bordalo, Pedro, Gennaioli, Nicola, and Shleifer, Andrei, “Salience and Consumer Choice,” *Journal of Political Economy*, 121 (2013), 803–843.
- [9] Bushong, Benjamin, Rabin, Matthew, and Schwartzstein, Joshua, “A Model of Relative Thinking,” working paper (2017).
- [10] Carandini, Matteo, and Heeger, David, “Summation and Division by Neurons in Primate Visual Cortex,” *Science*, 264 (1994), 1333–1335.
- [11] Carandini, Matteo, and Heeger, David, “Normalization as a Canonical Neural Computation,” *Nature Reviews: Neuroscience*, 13 (2012), 51–62.
- [12] Carpenter, Gregory, Glazer, Rashi, and Nakamoto, Kent, “Meaningful Brands From Meaningless Differentiation: The Dependence on Irrelevant Attributes,” *Journal of Marketing Research*, 31 (1994), 339–350.
- [13] Daviet, Remi, and Webb, Ryan, “Double Decoys and a Possible Parameterization: Empirical Analyses of Pairwise Normalization,” SSRN working paper (2019).
- [14] De Clippel, Geoffroy, and Eliaz, Kfir, “Reason-Based Choice: a Bargaining Rationale for the Attraction and Compromise Effects,” *Theoretical Economics*, 7 (2012), 125–162.
- [15] Doyle, John, O’Connor, David J, Reynolds, Gareth, and Bottomley, Paul, “The Robustness of the Asymmetrically Dominated Effect: Buying Frames, Phantom Alternatives, and In-Store Purchases,” *Psychology and Marketing*, 16 (1999), 225–243.
- [16] Frederick, Shane, Lee, Leonard, and Baskin, Ernest, “The Limits of Attraction,” *Journal of Marketing Research*, 51 (2014), 487–507.

- [17] Gans, Joshua, “On the Impossibility of Rational Choice under Incomplete Information,” *Journal of Economic Behavior and Organization*, 29 (1996), 287–309.
- [18] Geyskens, Inge, Gielens, Katrijn, and Gijsbrechts, Els, “Proliferating Private-Label Portfolios: How Introducing Economy and Premium Private Labels Influences Brand Choice,” *Journal of Marketing Research*, 47 (2010), 791–807.
- [19] Gourville, John, and Soman, Dilip, “Overchoice and Assortment Type: When and Why Variety Backfires,” *Marketing Science*, 24 (2005), 382–395.
- [20] Hamalainen, Raimo, and Alaja, Susanna, “The Threat of Weighting Biases in Environmental Decision Analysis,” *Ecological economics*, 68 (2008), 556–569.
- [21] Hartline, Keffer, and Ratliff, Floyd, “Inhibitory Interaction of Receptor Units in the Eye of Limulus,” *The Journal of General Physiology*, 40 (1957), 357–376.
- [22] Heeger, David, “Normalization of Cell Responses in Cat Striate Cortex,” *Visual Neuroscience*, 9 (1992), 181–197.
- [23] Holper, Lisa, van Brussel, Leo, Schmidt, Luzia, Schulthess, Sven, Burke, Christopher, Louie, Kenway, Seifritz, Erich, and Tobler, Philippe, “Adaptive Value Normalization in the Prefrontal Cortex Is Reduced by Memory Load,” *eNeuro: Cognition and Behavior*, 4 (2017), 1–20.
- [24] Hsee, Chrisopher, Loewenstein, George, Blount, Sally, and Bazerman, Max, “Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis,” *Psychological bulletin*, 125 (1999), 576–590.
- [25] Huber, Joel, Payne, John, and Puto, Christopher, “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research*, 9 (1982), 90–98.
- [26] Huber, Joel, Payne, John, and Puto, Christopher, “Let’s Be Honest About the Attraction Effect,” *Journal of Marketing Research*, 51 (2014), 520–525.
- [27] Humphrey, Steven, “Regret Aversion or Event-Splitting Effects? More Evidence Under Risk and Uncertainty,” *Journal of Risk and Uncertainty*, 11 (1995), 263–274.
- [28] Jacobi, Sarah, and Hobbs, Benjamin, “Quantifying and Mitigating the Splitting Bias and Other Value Tree-Induced Weighting Biases,” *Decision Analysis*, 4 (2007), 194–210.
- [29] Kahneman, Daniel, and Knetsch, Jack, “Valuing Public Goods: The Purchase of Moral Satisfaction,” *Journal of Environmental Economics and Management*, 22 (1992), 57–70.
- [30] Kahneman, Daniel, and Tversky, Amos, “Choices, Values, Frames,” *American Psychologist*, 39 (1984), 341–50.
- [31] Kamenica, Emir, “Contextual Inference in Markets: On the Informational Content of Product Lines,” *American Economic Review*, 98 (2008), 2127–2149.
- [32] Khaw, Mel, Glimcher, Paul, and Louie, Kenway, “Normalized Value Coding Explains Dynamic Adaptation in the Human Valuation Process,” *Proceedings of the National Academy of Sciences*, 114 (2017), 2127–2149.

- [33] Kivetz, Ran, Netzer, Oded, and Srinivasan, V, "Alternative Models for Capturing the Compromise Effect," *Journal of Marketing Research*, 41 (2004), 237–257.
- [34] Kivetz, Ran, Netzer, Oded, and Srinivasan, V, "Extending Compromise Effect Models to Complex Buying Situations and Other Context Effects," *Journal of Marketing Research*, 41 (2004), 262–268.
- [35] Koszegi, Botond, and Szeidl, Adam, "A Model of Focusing in Economic Choice," *Quarterly Journal of Economics*, 128 (2013), 53–104.
- [36] Lea, Amanda, and Ryan, Michael, "Irrationality In Mate Choice Revealed by Tungara Frogs," *Science*, 349 (2015), 964–966.
- [37] LoFaro, Thomas, Louie, Kenway, Webb, Ryan, and Glimcher, Paul, "The Temporal Dynamics of Cortical Normalization Models of Decision-Making," *Letters in Biomathematics*, 1 (2014), 209–220.
- [38] Louie, Kenway, Glimcher, Paul, and Webb, Ryan, "Adaptive Neural Coding: From Biological to Behavioral Decision-Making," *Current Opinion in Behavioral Sciences*, 5 (2015), 91–99.
- [39] Louie, Kenway, Grattan, Lauren, and Glimcher, Paul, "Reward Value-Based Gain Control: Divisive Normalization in Parietal Cortex," *Journal of Neuroscience*, 31 (2011), 10627–10639.
- [40] Louie, Kenway, LoFaro, Thomas, Webb, Ryan, and Glimcher, Paul, "Dynamic Divisive Normalization Predicts Time-Varying Value Coding in Decision-Related Circuits," *Journal of Neuroscience*, 34 (2014), 16046–16057.
- [41] Malkoc, Selin, Hedgcock, William, and Hoeffler, Steve, "Between a Rock and a Hard Place: The Failure of the Attraction Effect Among Unattractive Alternatives," *Journal of Consumer Psychology*, 23 (2013), 317–329.
- [42] Markman, Arthur, and Medin, Douglas, "Similarity and Alignment in Choice," *Organizational Behavior and Human Decision Processes*, 63 (1995), 117–130.
- [43] Marley, Anthony, "Context Dependent Probabilistic Choice Models Based on Measures of Binary Advantage," *Mathematical Social Sciences*, 21 (1991), 201–231.
- [44] Masatlioglu, Yusufcan, Nakajima, Daisuke, and Ozbay, Erkut, "Revealed Attention," *American Economic Review*, 102 (2012), 2183–2205.
- [45] May, Kenneth, "Intransitivity, Utility, and the Aggregation of Preference Patterns," *Econometrica*, 22 (1954), 1–13.
- [46] Mellers, Barbara, and Cooke, Alan, "Trade-Offs Depend On Attribute Range," *Journal of Experimental Psychology: Human Perception and Performance*, 20 (1994), 1055.
- [47] Meyer, Robert, Zhao, Shenghui, and Han, Jin, "Biases In Valuation Vs. Usage of Innovative Product Features," *Marketing Science*, 27 (2008), 1083–1096.
- [48] Mick, David, and Fournier, Susan, "Paradoxes Of Technology: Consumer Cognizance, Emotions, and Coping Strategies," *Journal of Consumer Research*, 25 (1998), 123–143.
- [49] Natenzon, Paulo, "Random Choice and Learning," forthcoming, *Journal of Political Economy* (2018).

- [50] Netzer, Nick, "Evolution of Time Preferences and Attitudes toward Risk," *American Economic Review*, 99 (2009), 937–955.
- [51] Noguchi, Takao, and Stewart, Neil, "In the Attraction, Compromise, and Similarity Effects, Alternatives Are Repeatedly Compared in Pairs on Single Dimensions," *Cognition*, 132 (2014), 44–56.
- [52] Noguchi, Takao, and Stewart, Neil, "Multialternative Decision By Sampling: A Model of Decision Making Constrained By Process Data," *Psychological Review*, 125 (2018), 512–544.
- [53] Ok, Efe, Ortoleva, Pietro, and Riella, Gil, "Revealed (P)Reference Theory," *American Economic Review*, 105 (2015), 299–321.
- [54] Rangel, Antonio, and Clithero, John, "Value Normalization in Decision Making: Theory and Evidence," *Current Opinion in Neurobiology*, 22 (2012), 970–981.
- [55] Rayo, Luis, and Becker, Gary, "Evolutionary Efficiency and Happiness," *Journal of Political Economy*, 115 (2007), 302–337.
- [56] Read, Daniel, and Loewenstein, George, "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices," *Journal of Experimental Psychology: Applied*, 1 (1995), 34.
- [57] Rieskamp, Jorg, Busemeyer, Jerome, and Mellers, Barbara, "Extending the Bounds of Rationality: Evidence and Theories of Preferential Choice," *Journal of Economic Literature*, 44 (2006), 631–661.
- [58] Robson, Arthur, "The Biological Basis of Economic Behavior," *Journal of Economic Literature*, 39 (2001), 11–33.
- [59] Robson, Arthur, and Whitehead, Lorne, "Rapidly Adaptive Hedonic Utility," working paper (2018).
- [60] Rubinstein, Ariel, "Irrational Diversification in Multiple Decision Problems," *European Economic Review*, 46 (2002), 1369–1378.
- [61] Russo, Edward, and Doshier, Barbara, "Strategies for Multiattribute Binary Choice," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9 (1983), 676–696.
- [62] Russo, Edward, and Rosen, Larry, "An Eye Fixation Analysis of Multialternative Choice," *Memory and Cognition*, 3 (1975), 267–276.
- [63] Schwartz, Janet, and Chapman, Gretchen, "Are More Options Always Better? The Attraction Effect in Physicians' Decisions About Medications," *Medical Decision Making*, 19 (1999), 315–323.
- [64] Shevell, Steven, "Saturation in Human Cones," *Vision Research*, 17 (1977), 427–434.
- [65] Shiell, Alan, and Gold, Lisa, "Contingent Valuation in Health Care and the Persistence of Embedding Effects without the Warm Glow," *Journal of Economic Psychology*, 23 (2002), 251–262.
- [66] Simonson, Itamar, "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16 (1989), 158–174.

- [67] Soltani, Alireza, De Martino, Benedetto, and Camerer, Colin, “A Range-Normalization Model of Context-Dependent Choice: A New Model and Evidence,” *PLoS Computational Biology*, 8 (2012), e1002607.
- [68] Starmer, Chris, and Sugden, Robert, “Testing for Juxtaposition and Event-Splitting Effects,” *Journal of Risk and Uncertainty*, 6 (1993), 235–254.
- [69] Steverson, Kai, Brandenburger, Adam, and Glimcher, Paul, “Rational Imprecision: Information-Processing, Neural, and Choice-Rule Perspectives,” working paper (2017).
- [70] Sullivan, Nicolette, Fitzsimons, Gavan, Platt, Michael, and Huettel, Scott, “Indulgent Foods Can Paradoxically Promote Disciplined Dietary Choices,” *Psychological Science*, 30 (2019), 273–287.
- [71] Thompson, Debora, Hamilton, Rebecca, and Rust, Roland, “Feature Fatigue: When Product Capabilities Become Too Much of a Good Thing,” *Journal of Marketing Research*, 42 (2005), 431–442.
- [72] Thompson, Debora, and Norton, Michael, “The Social Utility of Feature Creep,” *Journal of Marketing Research*, 48 (2011), 555–565.
- [73] Trueblood, Jennifer, Brown, Scott, Heathcote, Andrew, and Busemeyer, Jerome, “Not Just for Consumers: Context Effects are Fundamental to Decision Making,” *Psychological Science*, 24 (2013), 901–908.
- [74] Tserenjigmid, Gerelt, “Theory of Decisions by Intra-Dimensional Comparisons,” *Journal of Economic Theory*, 159 (2015), 326–338.
- [75] Tsetsos, Konstantinos, Moran, Rani, Moreland, James, Chater, Nick, Usher, Marius, and Summerfield, Christopher, “Economic Irrationality Is Optimal During Noisy Decision Making,” *Proceedings of the National Academy of Sciences*, 113 (2016), 3102–3107.
- [76] Tversky, Amos, “Elimination by Aspects: a Theory of Choice.,” *Psychological review*, 79 (1972), 281–299.
- [77] Tversky, Amos, and Russo, Edward, “Substitutability and Similarity in Binary Choices,” *Journal of Mathematical Psychology*, 6 (1969), 1–12.
- [78] Tversky, Amos, and Simonson, Itamar, “Context-Dependent Preferences,” *Management Science*, 39 (1993), 1179–1189.
- [79] Tymula, Agnieszka, and Glimcher, Paul, “Expected Subjective Value Theory: A Representation of Decision Under Risk and Certainty,” working paper (2016).
- [80] Webb, Ryan, Glimcher, Paul, and Louie, Kenway, “Rationalizing Context-Dependent Preferences: Divisive Normalization and Neurobiological Constraints on Choice,” working paper (2016).
- [81] Weber, Ernst, “On the Tactile Senses” (with translation of *De Tactu*), New York: Experimental Psychology Society (1834).
- [82] Weber, Martin, and Borchering, Katrin, “Behavioral Influences on Weight Judgments in Multiattribute Decision Making,” *European Journal of Operational Research*, 67 (1993), 1–12.

- [83] Weber, Martin, Eisenfuhr, Franz, and Von Winterfeldt, Detlof, "The Effects of Splitting Attributes on Weights in Multiattribute Utility Measurement," *Management Science*, 34 (1988), 431–445.
- [84] Wilson, Nathan, Runyan, Caroline, Wang, Forea, and Sur, Mriganka, "Division and Subtraction by Distinct Cortical Inhibitory Networks in Vivo," *Nature*, 488 (2012), 343–348.
- [85] Woodford, Michael, "Inattentive Valuation and Reference-Dependent Choice," working paper (2012).
- [86] Yeung, Catherine, and Soman, Dilip, "Attribute Evaluability and the Range Effect," *Journal of Consumer Research*, 32 (2005), 363–369.
- [87] Zhang, Jiao, Hsee, Christopher, and Xiao, Zhixing, "The Majority Rule in Individual Decision Making," *Organizational Behavior and Human Decision Processes*, 99 (2006), 102–111.
- [88] Zhang, Shi, and Markman, Arthur, "Overcoming the Early Entrant Advantage: The Role of Alignable and Nonalignable Differences," *Journal of Marketing Research*, 35 (1998), 413–426.



## A Additional Proofs

### A.1 Proof of Lemma 2

Using (1), we can see that  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  given Assumption BCI if and only if  $\frac{x_1}{x_1+z_1} + \frac{x_2}{x_2+z_2} > \frac{y_1}{y_1+z_1} + \frac{y_2}{y_2+z_2}$ , which we re-express as  $\frac{\tilde{x}_1}{\tilde{x}_1+\tilde{z}_1} + \frac{\tilde{x}_2}{\tilde{x}_2+\tilde{z}_2} > \frac{\tilde{y}_1}{\tilde{y}_1+\tilde{z}_1} + \frac{\tilde{y}_2}{\tilde{y}_2+\tilde{z}_2}$  with  $\tilde{w}_n \equiv \frac{w_n}{\sqrt{x_1x_2}}$  for  $\mathbf{w} \in \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  and  $n \in \{1, 2\}$ . Using  $\tilde{x}_1\tilde{x}_2 = \tilde{y}_1\tilde{y}_2 = 1$  to substitute out  $\tilde{x}_2$  and  $\tilde{y}_2$ , cross-multiplying and collecting terms, then factoring out  $\tilde{x}_1 - \tilde{y}_1 > 0$ , we can see this condition holds if and only if  $(1 - \tilde{z}_1\tilde{z}_2)(\tilde{z}_1 - \tilde{x}_1\tilde{y}_1\tilde{z}_2) > 0$ . Multiplying the first term by  $x_1x_2 > 0$  and the second by  $\frac{\sqrt{x_1x_2}}{z_2} > 0$ , then substituting out  $\frac{y_1}{x_2} = \frac{m_1^{xy}}{m_2^{xy}}$  (which holds since  $y_1m_2^{xy} = \frac{y_1x_2+y_1y_2}{2} = \frac{y_1x_2+x_1x_2}{2} = x_2m_1^{xy}$ ) along with each  $\tilde{w}_n = \frac{w_n}{\sqrt{x_1x_2}}$ , we can then see this is equivalent to  $(x_1x_2 - z_1z_2)(\frac{z_1}{z_2} - \frac{m_1^{xy}}{m_2^{xy}}) > 0$ . Noting  $\frac{z_1}{z_2} - \frac{m_1^{xy}}{m_2^{xy}} > 0$  since  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than  $\mathbf{y}$  and, from Lemma 1 and Proposition 13 (see Appendix C.4),  $x_1x_2 - z_1z_2 > 0$  holds if and only if  $\mathbf{x}$  is preferred to  $\mathbf{z}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  under Assumption BCI, this condition yields the desired result. ■

### A.2 Proof of Proposition 4

We proceed by contradiction. If there is a minority-rule preference cycle, then  $\mathbf{x}'$  is preferred to  $\mathbf{x}$ ,  $\mathbf{x}''$  to  $\mathbf{x}'$ , and  $\mathbf{x}$  to  $\mathbf{x}''$ . Let  $\lambda_1 \equiv \frac{1}{x'_1}$ ,  $\lambda_2 \equiv \frac{1}{x_2}$ ,  $\lambda_3 \equiv \frac{1}{x_3}$ , and  $\tilde{w}_n \equiv \lambda_n w_n$  for all  $\mathbf{w} \in \{\mathbf{x}, \mathbf{x}', \mathbf{x}''\}$  and  $n = 1, 2, 3$ . Also define  $k_n \equiv \max\{\tilde{w}_n\} - 1 > 0$  and  $q_n \equiv 1 - \min\{\tilde{w}_n\} > 0$  so that the ordered, rescaled attribute values are  $(1 + k_n, 1, 1 - q_n)$  for each  $n$ . Noting normalized valuations are invariant to scaling all attribute- $n$  values by  $\lambda_n > 0$ , our preference cycle implies:

$$\begin{aligned} V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) &\Rightarrow \frac{k_1}{2+k_1} + \frac{q_2}{2-q_2} < \frac{k_3+q_3}{2+k_3-q_3}, \\ V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) > V(\mathbf{x}''; \{\mathbf{x}, \mathbf{x}''\}) &\Rightarrow \frac{k_2}{2+k_2} + \frac{q_3}{2-q_3} < \frac{k_1+q_1}{2+k_1-q_1}, \\ V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\}) > V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) &\Rightarrow \frac{k_3}{2+k_3} + \frac{q_1}{2-q_1} < \frac{k_2+q_2}{2+k_2-q_2}. \end{aligned}$$

Summing these conditions yields  $\sum_{n=1}^3 (\frac{k_n}{2+k_n} + \frac{q_n}{2-q_n}) < \sum_{n=1}^3 (\frac{k_n+q_n}{2+k_n-q_n})$ . Thus,  $\frac{k_n}{2+k_n} + \frac{q_n}{2-q_n} < \frac{k_n+q_n}{2+k_n-q_n}$  for at least one  $n \in \{1, 2, 3\}$ . Combining the fractions on the left-side, we get  $\frac{2(k_n+q_n)}{(2+k_n)(2-q_n)} < \frac{k_n+q_n}{2+k_n-q_n}$ , which holds if and only if  $2(2+k_n-q_n) < (2+k_n)(2-q_n)$ , i.e. if and only if  $-q_nk_n > 0$ , a contradiction. ■

### A.3 Proof of Proposition 7

Follows from Proposition 7\* with  $R_n = 1$  for all  $n \leq N$  (see Appendix C.2). ■

#### A.4 Proof of Lemma 3.

Multiplying through by  $(\sigma + x_1 + x'_1)(\sigma + x_2 + x'_2) > 0$  and reducing the resulting expression, we see  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) > V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\})$  holds if and only if  $\sigma V^+(\mathbf{x}) + 2V^{\text{cd}}(\mathbf{x}) > \sigma V^+(\mathbf{x}') + 2V^{\text{cd}}(\mathbf{x}')$ . Parts (i) and (ii) are then readily verifiable from this inequality. ■

#### A.5 Proof of Proposition 9

*Part (i).* Using the notation in Proposition 1,  $\frac{\partial[V^*(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) - V^*(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\})]}{\partial k} = \frac{-2(y_2 - x_2)}{(\sigma + y_2 + x_2 + 2k)^2} < 0$  since  $y_2 > x_2$ , which ensures the relative difference effect holds under (3) for all  $\sigma \geq 0$ .

*Part (ii).* Let  $\hat{w}_n = w_n + \frac{\sigma}{2}$  for  $w \in \{x, x', x''\}$  and  $n = 1, 2, 3$ . Thus, a majority dominance relationship among  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  exists if and only if it also exists among  $\hat{\mathbf{x}}$ ,  $\hat{\mathbf{x}}'$ , and  $\hat{\mathbf{x}}''$ . From Proposition 4, if binary-choice preferences among  $\hat{\mathbf{x}}$ ,  $\hat{\mathbf{x}}'$ , and  $\hat{\mathbf{x}}''$  are intransitive, they must follow a majority-rule cycle in the basic PN model. Noting  $V(\hat{\mathbf{w}}; \{\hat{\mathbf{w}}, \hat{\mathbf{w}}'\}) = V^*(\mathbf{w}; \{\mathbf{w}, \mathbf{w}'\})$ , if preferences among  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$  are intransitive under (3) with  $\sigma \geq 0$ , they must also follow the same majority-rule cycle.

*Part (iii).* Using the notation in Proposition 5,  $\frac{x_{1a} - y_{1a}}{\sigma + x_1 + y_1} < \min\left\{\frac{x_{1a} - y_{1a}}{\sigma + x_{1a} + y_{1a}}, \frac{x_{1b} - y_{1b}}{\sigma + x_{1b} + y_{1b}}\right\}$  holds since  $x_1 + y_1 > \max\{x_{1a} + y_{1a}, x_{1b} + y_{1b}\}$ . Thus,  $\frac{x_{1a} - y_{1a}}{\sigma + x_1 + y_1} + \frac{x_{1b} - y_{1b}}{\sigma + x_{1b} + y_{1b}} > \frac{x_{1a} - y_{1a} + x_{1b} - y_{1b}}{\sigma + x_1 + y_1} = \frac{x_1 - y_1}{\sigma + x_1 + y_1}$ , which ensures the splitting bias holds under (3).

*Part (iv).* Using the notation in Proposition 6 and given  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V^*(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\})$ , we get  $V^*(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) - V^*(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\}) = V^*(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) - V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) - V^*(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\}) + V^*(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\}) = \frac{x_2}{\sigma + x_2} - \frac{y_2}{\sigma + y_2} - \frac{x_2 - y_2}{\sigma + x_2 + y_2} = \frac{(y_2 - x_2)x_2 y_2}{(\sigma + x_2)(\sigma + y_2)(\sigma + x_2 + y_2)} > 0$  for any  $\sigma \geq 0$  since  $y_2 > x_2$ . Thus, the alignability effect holds under (3).

*Part (v).* Using the notation in Proposition 7, we see  $V^*(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) - V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = \sum_{n \leq N} \frac{x'_n - x_n}{\sigma + x'_n + x_n} = \sum_{n < N} \frac{x'_n - AN^{-1}}{\sigma + x'_n + AN^{-1}} + \frac{A - \sum_{n < N} x'_n - AN^{-1}}{\sigma + A - \sum_{n < N} x'_n + AN^{-1}}$ . Differentiating by  $x'_n$ ,  $n \leq N - 1$  and substituting  $x'_N$  back in using  $\sum_{n \leq N} x'_n = A$  gives  $\frac{\sigma + 2AN^{-1}}{(\sigma + AN^{-1} + x'_n)^2} = \frac{\sigma + 2AN^{-1}}{(\sigma + AN^{-1} + x'_N)^2}$ . Thus, the system of  $N - 1$  first-order conditions is solved by  $x'_n = x'_N$ , implying  $x'_n = \frac{A}{N} = x_n$  for all  $n \leq N$ , thus ensuring  $V^*(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) < V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\})$  for  $\mathbf{x}' \neq \mathbf{x}$ .

*Part (vi).* Using the notation in Proposition 8,  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) - V^*(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = \frac{q}{\sigma + q} - \frac{q}{\sigma + 2x_{n'} + q} > 0$  since  $x_{n'} > 0$ . Thus, the feature bias holds for any  $\sigma \geq 0$  under (3). ■

#### A.6 Proof of Lemma 4

Using (1) and (3), we see  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) > V^*(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$  if and only if  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) > V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\})$ . Noting  $\frac{z'_1}{z'_2} = \frac{z_1 + \sigma}{z_2 + \sigma}$  and  $\frac{m_1^{\text{xy}}}{m_2^{\text{xy}}} = 1$  under Assumption BCI\*,  $\mathbf{z}'$  is more similar

to  $\mathbf{x}$  than to  $\mathbf{y}$  if and only if  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$ . From Lemmas 1 and 2, it then follows that  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) < V^*(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$  if and only if  $z_1 z_2 < x_1 x_2 < z'_1 z'_2$ . ■

### A.7 Proof of Proposition 10

Since  $x_1 x_2 > z_1 z_2$  must hold in the case of the compromise and dominance effects with  $\sigma = 0$ , the desired result then follows from Lemma 4. ■

### A.8 Proof of Corollary 1

Given  $V^*(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V^*(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$  under Assumption BCI\*, Proposition 10 implies  $V^{\text{CD}}(\mathbf{x}) = V^{\text{CD}}(\mathbf{z}')$ . For  $n = 1, 2$ , let  $\tilde{z}'_n = \tilde{z}_n + \sigma = \gamma z_n + \sigma$  and  $z'' = z_n + \frac{\sigma}{\gamma}$ . Noting  $\tilde{\mathbf{x}}$  and  $\tilde{\mathbf{y}}$  must also satisfy Assumption BCI\*, Proposition 10 also implies  $V^*(\tilde{\mathbf{x}}; \{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}\}) > V^*(\tilde{\mathbf{y}}; \{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}\})$  if and only if  $V^{\text{CD}}(\tilde{\mathbf{x}}) > V^{\text{CD}}(\tilde{\mathbf{z}}')$ . Since  $V^{\text{CD}}(\tilde{\mathbf{x}}) = \gamma^2 V^{\text{CD}}(\mathbf{x}) = \gamma^2 V^{\text{CD}}(\mathbf{z}')$ ,  $V^{\text{CD}}(\tilde{\mathbf{z}}) = \gamma^2 V^{\text{CD}}(\mathbf{z}'')$ , and  $V^{\text{CD}}(\mathbf{z}') \geq V^{\text{CD}}(\mathbf{z}'')$  if  $\gamma \geq 1$ ,  $\tilde{\mathbf{x}}$  will then be preferred to  $\tilde{\mathbf{y}}$  given  $X = \{\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}\}$  if  $\gamma > 1$ , while  $\tilde{\mathbf{y}}$  will be preferred to  $\tilde{\mathbf{x}}$  if  $\gamma < 1$ . ■

## B Additional Evidence of Majority-Rule Preference Cycles

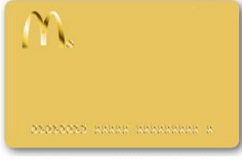








Some of the context-dependent behaviors captured by the basic PN model — most notably, majority-rule preference cycles, the splitting bias, the alignability effect, the diversification bias, and the feature bias (see Section 2) — are not explicitly addressed by prevailing multi-attribute choice theories. One may wonder if the lack of attention to these behaviors arises from concerns of empirical robustness. However, we are not aware of any research challenging the robustness of any of these effects, while each of these behaviors has been demonstrated in multiple empirical studies.

Even so, we still considered it worthwhile to test the robustness of majority-rule preference cycles due to their clear theoretical relevance — they directly contradict the canonical axiom that preferences are transitive — and also because the strongest existing evidence (at the time we decided to run the experiment) came from a rather dated, unincentivized study by May (1954).<sup>25</sup> While this study is not addressed by prevailing multi-attribute choice theories, it is worth noting that it has received some attention in the broader theoretical choice literature. In particular, Gans' (1996) 'small worlds' and Masatlioglu et al.'s (2012) 'limited consideration' theories both address May's (1954) findings.

In our experiment, 173 subjects (undergraduate students at the University of Toronto, Mississauga) were asked to make a sequence of binary choices among vacation packages

<sup>25</sup> Notably, we were not yet aware of Tsetsos et al.'s (2016) experiment, the results of which lend additional support to the notion that majority-rule preference cycles are a robust empirical phenomenon.

Figure 6: Example of Alternatives Used in Niagara Falls Choice Experiment

	Dining	Touring	Lodging
A	 <p><b>\$5 McDonald's Gift Card</b></p>	 <p><b>Bus Tour by Magnificent Tours</b> Includes: Journey Behind the Falls, Butterfly Conservatory, Niagara SkyWheel</p>	 <p><b>Marriott FallsView King Room</b> With Jacuzzi</p>
B	 <p><b>\$40 Ultimate Dining Gift Card*</b> can be used at any restaurant shown on card</p>	<p><b>Airbus H130 Helicopter Tour</b></p>  <p>"Memories to last a lifetime ... seeing the Falls from above gives you a perspective like no other"</p>	 <p><b>A1 Inn Motel</b> Driving Distance from Niagara Falls</p>
C	 <p><b>NIAGARA Sunset Cruises</b></p> <p>Dinner for two prepared by award-winning culinary team, served on a Niagara River paddle-wheeler cruise</p>	 <p>Option for <b>self-guided walking tour</b> using color city map printout</p>	 <p><b>Marriott Queen Room</b> No View</p>

to Niagara Falls. Each package was defined on three attribute dimensions: dining, touring, and lodging accommodations. The choice alternatives were constructed so that a majority-dominance relationship existed among them. The total cost of each option was approximately \$350 CAD. The choices were fully incentivized, with each subject entered into a lottery for which the winning subject received the package corresponding to one of their randomly-selected binary choices.

Of the 173 subjects, we found that 17, or 10% of subjects, displayed an intransitive cycle in their binary choices. This is a proportion of intransitivity typically observed in experimental data, though slightly lower than observed by May (1954). Of the 17 transitivity violations we observed, 13 of them (76.5%) were of the majority-rule form. This proportion is significantly greater than the proportion displaying the alternative minority-rule cycle ( $p = 0.044$ , two-tailed).

Certainly the basic PN model predicts a 100% incidence of majority-rule (and 0% minority-rule) violations when preferences are intransitive. With that said, it is well-known that choice is inherently random — an aspect that we abstract from in our deterministic model — and that randomness can generate intransitive choice. All else equal, however, majority- and minority-rule violations would be equally probable if stochasticity was the sole driver of intransitivity, yet a straightforward adaptation of our model (along the lines proposed in equation 4) would naturally predict that majority-rule preference cycles would

still be more probable than minority-rule preference cycles. Thus, pairwise normalization still provides a mechanism to explain evidence indicating a greater propensity for majority-rule preference cycles.

## C Additional Results

### C.1 Ease of Comparisons

In Section 4, we mentioned that pairwise normalization makes more similar alternatives “easier to compare” than less similar alternatives. Here, we formalize this idea:

**Proposition 11** *Under Assumption BCI, suppose the DM is not indifferent given  $X = \{\mathbf{x}, \mathbf{z}\}$  or  $X = \{\mathbf{y}, \mathbf{z}\}$ . Also suppose  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$  in that  $\frac{z_1}{z_2} > \frac{m_1^{xy}}{m_2^{xy}}$ . Then  $|V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})| > |V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\}) - V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\})|$ .*

*Proof.*  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})$  implies  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) \geq V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})$  from Lemma 2. Noting  $V(\mathbf{w}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{w}; \{\mathbf{x}, \mathbf{y}\}) + V(\mathbf{w}; \{\mathbf{x}, \mathbf{z}\})$  for  $\mathbf{w} \in \{\mathbf{x}, \mathbf{y}\}$  and  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\})$  under Assumption BCI, we see that  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})$  implies  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\})$ . This inequality, along with  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) = V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\}) + V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) = 2$ , implies  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) \leq V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\})$  for  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})$ , ensuring  $|V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})| > |V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\}) - V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\})|$ . ■

Proposition 11 indicates that the total perceived value difference will be larger when comparing  $\mathbf{z}$  to a more similar alternative  $\mathbf{x}$  than when comparing  $\mathbf{z}$  to a less similar alternative  $\mathbf{y}$ , despite  $\mathbf{x}$  and  $\mathbf{y}$  being equally valued when compared to each other. However, the implication that  $\mathbf{z}$  is easier to compare to  $\mathbf{x}$  than to  $\mathbf{y}$  in this deterministic sense is not directly observable as  $\mathbf{x}$  and  $\mathbf{y}$  would either both be preferred to (and hence, chosen over)  $\mathbf{z}$  with certainty in binary choice, or  $\mathbf{z}$  would be preferred to both. With this in mind, the following corollary shows how an adaptation of the basic PN model to a stochastic choice environment captures the ease of comparison concept in an observable form:

**Corollary 2** *Given the assumptions of Proposition 11, consider a stochastic extension of the deterministic basic PN model given in (1), with binary-choice probabilities given by*

$$\Pr[\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}] = f(V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}), V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\})), \quad (4)$$

*where  $f$  is strictly increasing in its first argument and strictly decreasing in its second argument. Then  $|\Pr[\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}] - \frac{1}{2}| < |\Pr[\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}] - \frac{1}{2}|$ .*

*Proof.* As shown in the proof of Proposition 11,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})$  implies  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\})$  and  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) \leq V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\})$ . Noting  $\Pr[\mathbf{z}; \{\mathbf{w}, \mathbf{z}\}] = f(V(\mathbf{z}; \{\mathbf{w}, \mathbf{z}\}), V(\mathbf{w}; \{\mathbf{w}, \mathbf{z}\}))$  for  $\mathbf{w} \in \{\mathbf{x}, \mathbf{y}\}$ ,  $\Pr[\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}] \leq \Pr[\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}]$  is assured for

$V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) \geq V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\})$  since  $f$  is increasing in its first argument and decreasing in its second argument. Thus, either  $\Pr[\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}] < \Pr[\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}] < \frac{1}{2}$  or  $\Pr[\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}] > \Pr[\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}] > \frac{1}{2}$  must hold, ensuring  $|\Pr[\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}] - \frac{1}{2}| < |\Pr[\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}] - \frac{1}{2}|$ . ■

Corollary 2 can be understood as follows. Suppose  $\mathbf{x}$  and  $\mathbf{y}$  are equally likely to be chosen in binary choice (as indirectly implied by (4)). Also suppose that the probabilities of choosing  $\mathbf{z}$  in binary choices with  $\mathbf{x}$  and with  $\mathbf{y}$  are both less than one half, suggesting  $\mathbf{z}$  is inferior to  $\mathbf{x}$  and  $\mathbf{y}$ . Then, if  $\mathbf{z}$  is more similar to  $\mathbf{x}$  than to  $\mathbf{y}$ , the likelihood of choosing  $\mathbf{z}$  in a binary choice is lower with  $X = \{\mathbf{x}, \mathbf{z}\}$  than with  $X = \{\mathbf{y}, \mathbf{z}\}$ . That is,  $\mathbf{z}$  is easier to compare to the similar alternative  $\mathbf{x}$  than to the less similar alternative  $\mathbf{y}$  in the sense that there is a lower probability that the DM will “mistakenly” choose the inferior alternative  $\mathbf{z}$  with  $\mathbf{x}$  than with  $\mathbf{y}$ .

## C.2 Allocation and Investment Results with Unequal Returns

As mentioned in Section 6, the results capturing the diversification and feature biases still hold in the basic PN model even if the returns to allocations along each dimension are not equal. We will now formalize and prove these results.

To begin, we now distinguish between the amount of an allocation to a given dimension and the attribute value generated by that allocation. In particular, we now let  $a_n$  denote the allocation of  $A > 0$  to dimension  $n \leq N$ , where the (unnormalized) attribute value generated by this allocation is now  $R_n a_n$  given  $R_n > 0$  is the (gross) rate of return on dimension  $n$ . In the following generalization of Proposition 7, we will assume that  $\mathbf{x}$  and  $\mathbf{x}'$  are the alternatives associated with the allocations  $a_1, \dots, a_N$  and  $a'_1, \dots, a'_N$ , respectively (with  $\sum_{n \leq N} a_n = \sum_{n \leq N} a'_n = A$ ), implying  $x_n = R_n a_n$  and  $x'_n = R_n a'_n$  for all  $n \leq N$ .

**Proposition 7\*** *Given  $N > 1$ ,  $A > 0$ , and  $R_n > 0$  for all  $n \leq N$ , suppose  $x_n = R_n a_n$  with  $a_n = \frac{A}{N}$  for all  $n \leq N$ . Then, for any  $\mathbf{x}' \neq \mathbf{x}$  satisfying  $x'_n = R_n a'_n$  with  $\sum_{n \leq N} a'_n = A$ ,  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ .*

*Proof.* Using  $\sum_{n \leq N} a'_n = A$  and  $x_n = R_n \frac{A}{N}$  to substitute out  $x'_N = R_N a'_N$  and each  $x_n$  from (1) while canceling all  $R_n$  terms gives  $V(\mathbf{x}'; X) = \sum_{n=1}^{N-1} \frac{a'_n}{A N^{-1} + a'_n} + \frac{A - \sum_{n=1}^{N-1} a'_n}{A N^{-1} + A - \sum_{n=1}^{N-1} a'_n}$ . Differentiating by  $a'_n$ ,  $n \leq N - 1$ , setting each derivative equal to zero, and substituting  $a'_N$  back in using  $\sum_{n \leq N} a'_n = A$  gives  $\frac{A N}{(A + a'_N N)^2} = \frac{A N}{(A + a'_N N)^2}$ . Thus, the system of  $N - 1$  first-order conditions is solved by  $a'_n = a'_N$ , implying  $a'_n = \frac{A}{N}$  (and  $x'_n = x_n$ ) for all  $n \leq N$ , ensuring  $V(\mathbf{x}'; X) < V(\mathbf{x}; X)$  for  $\mathbf{x}' \neq \mathbf{x}$ . ■

Thus, the diversification bias as captured in Proposition 7 still holds when we allow the rates of return to vary across dimensions.

To formalize the feature bias with unequal returns, we now assume that an investment of  $q > 0$  on dimension  $n \leq N$  yields a  $R_n q$  increase in the unnormalized attribute value on dimension  $n$ . We can then generalize Proposition 8 as:

**Proposition 8\*** *Given  $N > 1$ ,  $q > 0$ , and  $R_n > 0$  for  $n = 1, \dots, N$ , suppose  $x_N = R_N \cdot q$ ,  $x'_N = 0$ ,  $x'_{n'} = x_{n'} + R_{n'} \cdot q$  for some  $n' < N$ , and  $x'_n = x_n > 0$  for all  $n < N$ . Then  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ .*

*Proof.*  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) - V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = \Delta(R_N \cdot q, 0) - \Delta(x_{n'} + R_{n'} \cdot q, x_{n'}) = \frac{2x_{n'}}{R_{n'} \cdot q + 2x_{n'}} > 0$  given  $q > 0$ ,  $R_{n'} > 0$ , and  $x_{n'} > 0$ . ■

Thus, the feature bias as captured in Proposition 8 also still holds when we allow the rates of return to vary across dimensions.

### C.3 Range-Dependent Preferences

The following result shows how the perceived value difference between two attribute values decreases with the range of values on that dimension (holding the average fixed):

**Proposition 12** *Suppose the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\}$ , with  $x'_2 < x_2 < y_2 < y'_2$ . Also suppose  $x''_1 = x'_1$ ,  $y''_1 = y'_1$ ,  $x''_2 = x'_2 - k$ , and  $y''_2 = y'_2 + k$  for some  $k > 0$ . Then  $\mathbf{x}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}$ .*

*Proof.* Using  $x''_1 = x'_1$ ,  $y''_1 = y'_1$ ,  $x''_2 = x'_2 - k$ , and  $y''_2 = y'_2 + k$ , we can express  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}) - V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\}) = \frac{z_2}{z_2 + x'_2 - k} + \frac{z_2}{z_2 + y'_2 + k} - \frac{z_2}{z_2 + x'_2} - \frac{z_2}{z_2 + y'_2}$  for each  $z_2 \in \{x_2, y_2\}$ . Hence,  $\frac{\partial^2}{\partial z_2 \partial k} [V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}) - V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\})]_{k=0} = \frac{x'_2 - z_2}{(z_2 + x'_2)^3} - \frac{y'_2 - z_2}{(z_2 + y'_2)^3} < 0$  given  $x'_2 < z_2 < y'_2$ . Thus,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}) - V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\}) > V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}) - V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\})$  since  $y_2 > x_2$ , implying  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\}) > V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}'', \mathbf{y}''\})$  given  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}'\})$ . ■

While Proposition 1 demonstrated how an increase in the average attribute value shifted preferences in favor of the alternative that was weaker on that dimension, Proposition 12 demonstrates how an increase in the range of attribute values has the same effect, in line with evidence from Mellers and Cooke (1994) and Yeung and Soman (2005).

### C.4 Superiority/Inferiority Result

As mentioned in footnote 15, the following result shows that, when the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice, the superiority or inferiority of  $\mathbf{z}$  relative to  $\mathbf{x}$  and  $\mathbf{y}$  is maintained in trinary choice.

**Proposition 13** *Under Assumption BCI, the following are equivalent:*

- (i-a)  $\mathbf{z}$  is preferred to  $\mathbf{x}$  given  $X = \{\mathbf{x}, \mathbf{z}\}$ ,
- (i-b)  $\mathbf{z}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{y}, \mathbf{z}\}$ ,
- (ii-a)  $\mathbf{z}$  is preferred to  $\mathbf{x}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,
- (ii-b)  $\mathbf{z}$  is preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ .

*Proof.* From Lemma 1,  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  (i-a) is equivalent to  $z_1 z_2 > x_1 x_2$  while Assumption BCI implies  $x_1 x_2 = y_1 y_2$ . Taken together, (i-a) must be equivalent to  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\})$  (i-b). In turn, we can see  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) + V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\}) = \frac{x'_1 + x''_1}{x'_1 + x''_1} + \frac{x'_2 + x''_2}{x'_2 + x''_2} = 2$ . Therefore,  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) > V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\})$ ,  $1 > V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\})$ , and  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) > 1$ , are all equivalent. Since  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  and  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\})$  are equivalent (seen above),  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  implies  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) + 1$ . Since  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\})$ ,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = 1$ . Thus,  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  implies  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{x}; \{\mathbf{y}, \mathbf{z}\})$  (ii-a). In turn, we can also see  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\}) + V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{x}; \{\mathbf{y}, \mathbf{z}\})$  implies  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  or  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = 1$  or both. Since  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > 1$  is equivalent to  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{y}; \{\mathbf{y}, \mathbf{z}\})$ , which itself is equivalent to  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$ , since at least one among  $V(\mathbf{z}; \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{z}\})$  and  $V(\mathbf{z}; \{\mathbf{y}, \mathbf{z}\}) > V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = 1$  are true, they must both hold. This establishes the equivalence of (ii-a) and (i-a). By switching  $\mathbf{x}$  and  $\mathbf{y}$  in the arguments outlined above, we can likewise establish the equivalence of (ii-b) and (i-b). ■

## C.5 Two-Parameter Model Results

Next, we provide additional results arising from a two-parameter version of the pairwise normalization model (based on a common formulation of the normalization computation studied in neuroscience). Given  $\sigma \geq 0$  and  $\alpha > 0$ , the two-parameter model is given by:

$$V^{**}(\mathbf{x}; X) = \sum_{n=1}^N \sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{x_n^\alpha}{\sigma^\alpha + x_n^\alpha + y_n^\alpha}. \quad (5)$$

The following result shows how the two-parameter model nests some classic microeconomic preference representations when applied to two-attribute binary choice problems:

**Proposition 14** *Given  $X = \{\mathbf{x}, \mathbf{x}'\}$  and  $N = 2$  under (5). For each of the following specifications of  $\tilde{V}(a, b)$  with the indicated parametric restrictions,  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  if  $\tilde{V}(\mathbf{x}) > \tilde{V}(\mathbf{x}')$ :*

- (i)  $\tilde{V}(a, b) = V^{\text{CD}}(a, b) \equiv ab$ , with  $\sigma = 0$  and any  $\alpha > 0$ .
- (ii)  $\tilde{V}(a, b) = V^{\text{CES}}(a, b) \equiv (a^\alpha + b^\alpha)^{1/\alpha}$ , with  $\sigma > 0$  sufficiently large and any  $\alpha > 0$ .



(iii)  $\tilde{V}(a, b) = \max\{a, b\}$ , with  $\sigma > 0$  and  $\alpha > 0$  both sufficiently large; if  $\tilde{V}(\mathbf{x}) = \tilde{V}(\mathbf{x}')$ ,  $\mathbf{x}$  is then preferred to  $\mathbf{x}'$  if and only if  $\tilde{V}_0(\mathbf{x}) > \tilde{V}_0(\mathbf{x}')$ , where  $\tilde{V}_0(a, b) = \min\{a, b\}$ .

*Proof.* Given  $X = \{\mathbf{x}, \mathbf{x}'\}$ ,  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  if and only if  $V^{**}(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) - V^{**}(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = \sum_n \frac{x_n^\alpha - x_n'^\alpha}{\sigma^\alpha + x_n^\alpha + x_n'^\alpha} > 0$ . Combining terms, factoring out the denominator, and taking  $N = 2$  yields:

$$\sigma^\alpha(x_1^\alpha + x_2^\alpha) + 2x_1^\alpha x_2^\alpha > \sigma^\alpha(x_1'^\alpha + x_2'^\alpha) + 2x_1'^\alpha x_2'^\alpha, \quad (6)$$

so that  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  given  $\sigma = 0$  if and only if  $x_1^\alpha x_2^\alpha > x_1'^\alpha x_2'^\alpha$ , which is equivalent to  $x_1 x_2 > x_1' x_2'$ . This establishes part (i).

For part (ii),  $\tilde{V}(\mathbf{x}) > \tilde{V}(\mathbf{x}')$  if and only if  $\tilde{V}(\mathbf{x})^\alpha > \tilde{V}(\mathbf{x}')^\alpha$ , which is equivalent to  $x_1^\alpha + x_2^\alpha > x_1'^\alpha + x_2'^\alpha$  given  $\tilde{V}(a, b) = (a^\alpha + b^\alpha)^{1/\alpha}$ . Let  $\sigma_0 = \left( \frac{2(y_1^\alpha y_2^\alpha - x_1^\alpha x_2^\alpha)}{x_1^\alpha + x_2^\alpha - x_1'^\alpha - x_2'^\alpha} \right)^{1/\alpha} < \infty$ . Observe  $\sigma_0^\alpha(x_1^\alpha + x_2^\alpha) + 2x_1^\alpha x_2^\alpha = \sigma_0^\alpha(x_1'^\alpha + x_2'^\alpha) + 2x_1'^\alpha x_2'^\alpha$ . Thus,  $\sigma^\alpha(x_1^\alpha + x_2^\alpha) + 2x_1^\alpha x_2^\alpha > \sigma^\alpha(x_1'^\alpha + x_2'^\alpha) + 2x_1'^\alpha x_2'^\alpha$  for all  $\sigma > \sigma_0$ , implying  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  from (6). The converse is established by contradiction. Namely, suppose  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  but  $\tilde{V}(\mathbf{x}) < \tilde{V}(\mathbf{x}')$ , or equivalently,  $x_1^\alpha + x_2^\alpha < x_1'^\alpha + x_2'^\alpha$ . From (6), we see that, together, these conditions require  $x_1^\alpha x_2^\alpha > x_1'^\alpha x_2'^\alpha$ , so that  $x_1^\alpha x_2^\alpha - x_1'^\alpha x_2'^\alpha > 0$ . By inspection, we can now see  $\sigma > \sigma_0$  with  $\sigma_0 > 0$  as defined above implies  $\sigma^\alpha(x_1^\alpha + x_2^\alpha) + 2x_1^\alpha x_2^\alpha < \sigma^\alpha(x_1'^\alpha + x_2'^\alpha) + 2x_1'^\alpha x_2'^\alpha$ , which from (6) implies  $\mathbf{x}'$  is preferred to  $\mathbf{x}$ . Hence, we have a contradiction, so that a preference for  $\mathbf{x}$  over  $\mathbf{x}'$  necessarily requires  $\tilde{V}(\mathbf{x}) > \tilde{V}(\mathbf{x}')$  for sufficiently large  $\sigma > 0$ , as desired.

For part (iii), given  $\tilde{V}(a, b) = \max\{a, b\}$ , letting  $x = \max\{x_1, x_2\}$  and  $x' = \max\{x_1', x_2'\}$ , without loss of generality, we see  $\tilde{V}(\mathbf{x}) > \tilde{V}(\mathbf{x}')$  holds if and only if  $x > x'$ . Observe,  $\sigma^\alpha(x_1^\alpha + x_2^\alpha) + 2x_1^\alpha x_2^\alpha \geq \sigma^\alpha x^\alpha$ . Given any  $\sigma > x'$ , we also see  $\sigma^\alpha(x_1'^\alpha + x_2'^\alpha) + 2x_1'^\alpha x_2'^\alpha \leq 2\sigma^\alpha x_1'^\alpha + 2x_1'^{2\alpha} < 4\sigma^\alpha x'^\alpha$ . From (6), we can then see that a sufficient condition for  $\mathbf{x}$  to be preferred to  $\mathbf{x}'$  given any  $\sigma > x'$  is  $\sigma^\alpha x^\alpha > 4\sigma^\alpha x'^\alpha$ . Factoring out  $\sigma^\alpha > 0$  then taking the natural log, we see this condition is equivalent to  $\alpha \ln(x) > \alpha \ln(x') + \ln(4)$ . Taking  $\alpha_0 \equiv \frac{\ln(4)}{\ln(x) - \ln(x')} > 0$ , we see  $\alpha \ln(x) > \alpha \ln(x') + \ln(4)$  holds for any  $\alpha > \alpha_0$  and  $\sigma > x'$ , so that  $\mathbf{x}$  must be preferred to  $\mathbf{x}'$  for sufficiently large  $\alpha$  and  $\sigma$ . The converse is established by contradiction. Suppose  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  but  $\tilde{V}(\mathbf{x}) < \tilde{V}(\mathbf{x}')$ , or equivalently,  $x' > x$ . Using (6) while applying the logic outlined above (except switching the roles of  $\mathbf{x}$  and  $\mathbf{x}'$ ), it must be the case that, for any  $\sigma > x$ ,  $\alpha \ln(x) + \ln(4) > \alpha \ln(x')$  by virtue of the preference for  $\mathbf{x}$  over  $\mathbf{x}'$ . Defining  $\alpha'_0 \equiv \frac{\ln(4)}{\ln(x') - \ln(x)} > 0$  (positive because  $x' > x$ ), we can see  $\alpha > \alpha'_0$  implies  $\alpha \ln(x) + \ln(4) < \alpha \ln(x')$ . Hence, we have a contradiction, so that a preference for  $\mathbf{x}$  over  $\mathbf{x}'$  (with  $\tilde{V}(\mathbf{x}) \neq \tilde{V}(\mathbf{x}')$ ) must require  $\tilde{V}(\mathbf{x}) > \tilde{V}(\mathbf{x}')$  for sufficiently large  $\sigma > 0$  and  $\alpha > 0$ . In the case of  $\tilde{V}(\mathbf{x}) = \tilde{V}(\mathbf{x}')$ , i.e.,  $x = x'$ , we see from (6) that, in this case,  $\mathbf{x}$  will be preferred to  $\mathbf{x}'$  if and only if  $\sigma^\alpha(x_+^\alpha + x_-^\alpha) + 2x_+^\alpha x_-^\alpha > \sigma^\alpha(x_+'^\alpha + x_-'^\alpha) + 2x_+'^\alpha x_-'^\alpha$  with  $x_- \equiv \min\{x_1, x_2\}$  and  $x'_- \equiv \min\{x_1', x_2'\}$ . Subtracting  $\sigma^\alpha x_+^\alpha$  from both sides, then factoring out  $\sigma^\alpha + 2x_+^\alpha > 0$ , we see this is equivalent to  $x_- > x'_-$ . Given  $\tilde{V}_0(a, b) \equiv \min\{a, b\}$  with  $x \geq x_-$  and  $x' \geq x'_-$ , we see that  $x_- > x'_-$  is equivalent to  $\tilde{V}_0(\mathbf{x}) > \tilde{V}_0(\mathbf{x}')$ . ■

Part (i) of Proposition 14 shows that the previously-established equivalence between the basic PN model and the (symmetric) Cobb-Douglas preference model (Lemma 1) extends to any  $\alpha > 0$ , provided  $\sigma = 0$  is maintained. Part (ii) shows that preferences converge to those represented by a constant elasticity of substitution (CES) preference model in the large- $\sigma$  limit of the two-parameter PN model. In this case,  $(1 - \alpha)^{-1}$  represents the effective elasticity of substitution across attributes, implying preferences are nonconvex if  $\alpha > 1$  (i.e. if  $(1 - \alpha)^{-1} < 0$ ). Lastly, part (iii) shows that when  $\sigma$  and  $\alpha$  are both arbitrarily large, preferences align with those given by a rank-based lexicographic model, in which the preference between  $\mathbf{x}$  and  $\mathbf{x}'$  is determined by each alternative's larger attribute value ( $\max\{x_1, x_2\}$ ,  $\max\{x'_1, x'_2\}$ ). In the event of a tie, the preference is then determined by their smaller attribute values ( $\min\{x_1, x_2\}$ ,  $\min\{x'_1, x'_2\}$ ).

More generally, binary-choice preferences among two-attribute alternatives in the two-parameter model are a composite of preferences under the Cobb-Douglas and CES preference models, with  $\sigma$  determining the relative weight of each representation:

**Proposition 15** *Given  $N = 2$  and  $X = \{\mathbf{x}, \mathbf{x}'\}$  under (5):*

- (i) *If  $V^{\text{CD}}(\mathbf{x}) \geq V^{\text{CD}}(\mathbf{x}')$ ,  $V^{\text{CES}}(\mathbf{x}) \geq V^{\text{CES}}(\mathbf{x}')$ , and  $\sigma \geq 0$  with at least two of these inequalities non-binding, then  $\mathbf{x}$  is preferred to  $\mathbf{x}'$ .*
- (ii) *If  $V^{\text{CD}}(\mathbf{x}) > V^{\text{CD}}(\mathbf{x}')$  and  $V^{\text{CES}}(\mathbf{x}') > V^{\text{CES}}(\mathbf{x})$ , there exists a  $\sigma_0 > 0$  (determined by  $x_1, x_2, x'_1, x'_2$ , and  $\alpha$ ) for which  $\mathbf{x}$  is preferred to  $\mathbf{x}'$  if and only if  $\sigma < \sigma_0$ .*

*Proof.* Using (6), we can see that  $\mathbf{x}$  will be preferred to  $\mathbf{x}'$  if and only if  $(\sigma V^{\text{CES}}(\mathbf{x}))^\alpha + 2(V^{\text{CD}}(\mathbf{x}))^\alpha > (\sigma V^{\text{CES}}(\mathbf{x}'))^\alpha + 2(V^{\text{CD}}(\mathbf{x}'))^\alpha$ , from which the result in part (i) is readily verified. Part (ii) is also readily verifiable from the condition for  $\mathbf{x}$  to be preferred to  $\mathbf{x}'$ , as re-expressed in part (i), where  $\sigma_0 \equiv \left( \frac{2((V^{\text{CD}}(\mathbf{x}))^\alpha - (V^{\text{CD}}(\mathbf{x}'))^\alpha)}{(V^{\text{CES}}(\mathbf{x}'))^\alpha - (V^{\text{CES}}(\mathbf{x}))^\alpha} \right)^{1/\alpha}$  is derived from the implied indifference condition. ■

Thus, if the Cobb-Douglas and CES preference models agree in their rankings among the two alternatives, the DM's preference will align with this ranking. If they disagree, preferences will coincide with Cobb-Douglas if  $\sigma < \sigma_0$  and with CES if  $\sigma > \sigma_0$ , for some threshold  $\sigma_0 > 0$  — in effect, a higher  $\sigma$  implies a larger weight of CES relative to Cobb-Douglas preferences in determining the DM's preference.

The next result shows that, unlike the contrast function  $\Delta$  from the basic PN model, the analogous contrast function under (5), denoted as  $\Delta^{**}$ , does not exhibit diminishing sensitivity over its full domain:

**Proposition 16** *Given  $x_n \geq y_n$  (without loss of generality),  $\hat{\sigma}(y_n) \equiv \left(\frac{2}{\alpha-1}\right)^{1/\alpha} y_n$ , and  $\Delta^{**}(x_n, y_n) \equiv \left| \frac{x_n^\alpha - y_n^\alpha}{\sigma^\alpha + x_n^\alpha + y_n^\alpha} \right|$ :*

- (i)  *$\Delta^{**}(x_n, y_n)$  satisfies diminishing sensitivity if and only if  $\sigma = 0$  or  $\alpha \leq 1$  (or both).*

(ii) If  $\alpha > 1$  and  $\sigma \leq \hat{\sigma}(y_n)$ ,  $\Delta^{**}(x_n, y_n)$  exhibits diminishing sensitivity (locally) and is concave in  $x_n$  for all  $x_n \geq y_n$ .

(iii) If  $\alpha > 1$  and  $\sigma > \hat{\sigma}(y_n)$ , there exist increasing functions  $\bar{x}(\sigma) > y_n$  and  $\hat{x}(\sigma) > y_n$  such that  $\Delta^{**}(x_n, y_n)$  exhibits diminishing sensitivity if and only if  $x_n > \bar{x}(\sigma)$ , and is concave in  $x_n$  if and only if  $x_n > \hat{x}(\sigma)$ .

*Proof.* For part (i), note  $\frac{d[\Delta^{**}(x_n + \epsilon, y_n + \epsilon)]}{d\epsilon} = -\frac{\alpha x_n^\alpha y_n^\alpha (2(x_n^\alpha - y_n^\alpha) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha}))}{x_n y_n (x_n^\alpha + y_n^\alpha + \sigma^\alpha)^2}$  given  $x_n \geq y_n$  (without loss of generality). Thus,  $\frac{d[\Delta^{**}(x_n + \epsilon, y_n + \epsilon)]}{d\epsilon} < 0$  if and only if  $2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha}) > 0$ . With  $x_n \geq y_n$ , this clearly holds for  $\sigma = 0$  and also for  $\alpha \leq 1$  because, together,  $\alpha \leq 1$  and  $x_n \geq y_n$  guarantee  $x_n^{1-\alpha} - y_n^{1-\alpha} \geq 0$ . Thus,  $\Delta^{**}(x_n + \epsilon, y_n + \epsilon) < \Delta^{**}(x_n, y_n)$  for all  $\epsilon > 0$  given  $\sigma = 0$  or  $\alpha \leq 1$  (or both). Thus, to complete the proof, we only need to show that for any  $\sigma > 0$  and  $\alpha > 1$ , there exist a  $x_n \geq 0$  and  $y_n \geq 0$  with  $x_n \geq y_n$  such that  $2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha}) < 0$ . Take  $y_n = \frac{\sigma(\alpha-1)^{1/\alpha}}{2}$  and let  $x_n = y_n + \delta$ . Substituting these into  $2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha})$  then differentiating with respect to  $\delta$ , we get  $2 - 2^\alpha < 0$  for  $\alpha > 1$ . Also note  $2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha}) = 0$  given  $x_n = y_n$ , i.e., given  $\delta = 0$ . Together, these imply  $2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha}) < 0$  for  $y_n = \frac{\sigma(\alpha-1)^{1/\alpha}}{2}$  and  $x_n = y_n + \delta$ , provided  $\delta > 0$  is sufficiently small, as desired.

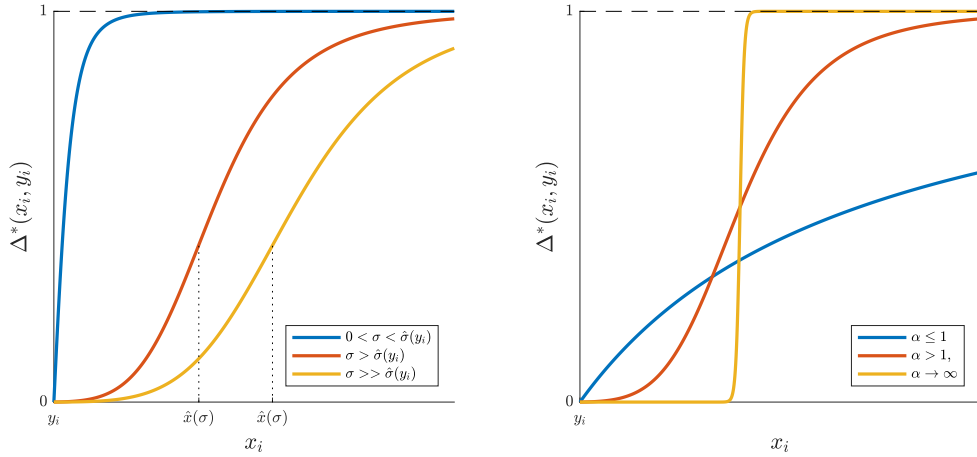
For part (ii), let  $h(x_n|y_n, \sigma, \alpha) \equiv 2(x_n - y_n) + \sigma^\alpha (x_n^{1-\alpha} - y_n^{1-\alpha})$ . From our above work, we can see that  $\Delta^{**}(x_n, y_n)$  exhibits diminishing sensitivity for all  $x_n \geq y_n$  if and only if  $h(x_n|y_n, \sigma, \alpha) > 0$  for all  $x_n \geq y_n$ . Note  $h'(x_n|y_n, \sigma, \alpha) = 2 - \frac{(\alpha-1)\sigma^\alpha}{x_n^\alpha}$ , so that  $h'(x_n|y_n, \sigma, \alpha) = 0$  if and only if  $x_n = x_n^* \equiv \sigma \left(\frac{\alpha-1}{2}\right)^{1/\alpha}$ . Also note,  $h''(x_n|y_n, \sigma, \alpha) = \frac{\alpha(\alpha-1)\sigma^\alpha}{x_n^{\alpha+1}} > 0$ . Thus,  $x_n = x_n^*$  uniquely maximizes  $h(x_n|y_n, \sigma, \alpha)$ . Given  $h'(x_n|y_n, \sigma, \alpha) > 0$  for all  $x_n > x_n^*$  and  $h(y_n|y_n, \sigma, \alpha) = 0$ ,  $\Delta^{**}(x_n, y_n)$  satisfies diminishing sensitivity for all  $x_n \geq y_n$  if and only if  $x_n^* \leq y_n$  or  $h(x_n^*|y_n, \sigma, \alpha) \geq 0$  (or both). Given  $h'(x_n|y_n, \sigma, \alpha) < 0$  for all  $x_n < x_n^*$  and  $h(y_n|y_n, \sigma, \alpha) = 0$ ,  $x_n^* > y_n$  implies  $h(x_n^*|y_n, \sigma, \alpha) < 0$ . Taken together, these last two observations imply  $\Delta^{**}(x_n, y_n)$  satisfies diminishing sensitivity for all  $x_n \geq y_n$  if and only if  $x_n^* \leq y_n$ , which, using the definitions of  $x_n^*$  and of  $\hat{\sigma}(y_n)$ , we can see this is equivalent to  $\hat{\sigma}(y_n) = \left(\frac{2}{\alpha-1}\right)^{1/\alpha} y_n$ . Computing  $\frac{\partial^2 \Delta^{**}(x_n, y_n)}{\partial x_n^2}$ , multiplying through by  $x_n^2(x_n^\alpha + y_n^\alpha + \sigma^\alpha) > 0$ , dividing by  $\alpha x_n^\alpha (2y_n^\alpha + \sigma^\alpha) > 0$ , and rearranging, we see  $\Delta^{**}(x_n, y_n)$  is concave in  $x_n$  for all  $x_n \geq y_n$  if and only if  $x_n^\alpha (1 + \alpha) \geq (\alpha - 1)(y_n^\alpha + \sigma^\alpha)$  for all  $x_n \geq y_n$ . Since the left-side of this inequality is clearly increasing in  $x_n$ ,  $\Delta^{**}(x_n, y_n)$  is concave in  $x_n$  for all  $x_n \geq y_n$  if and only if the inequality holds at  $x_n = y_n$ , i.e., if and only if  $y_n^\alpha (1 + \alpha) \geq (\alpha - 1)(y_n^\alpha + \sigma^\alpha)$ . Solving for  $\sigma$ , we see this condition is equivalent to  $\sigma \leq \hat{\sigma}(y_n) = \left(\frac{2}{\alpha-1}\right)^{1/\alpha} y_n$ , as desired.

For part (iii), let  $\bar{x}(\sigma) \equiv \{x_n : \sigma^\alpha = \frac{2(x_n - y_n)}{x_n^{1-\alpha} - y_n^{1-\alpha}}\} > y_n$ , and  $\hat{x}(\sigma) \equiv \left(\frac{(\alpha-1)(y_n^\alpha + \sigma^\alpha)}{\alpha+1}\right)^{1/\alpha} > y_n$ . Using our definitions of  $h(x_n|y_n, \sigma, \alpha)$  and  $\bar{x}(\sigma)$ ,  $h(\bar{x}(\sigma)|y_n, \sigma, \alpha) = 0$  is readily verifiable. Given  $x_n^* > y_n$  for  $\sigma > \hat{\sigma}(y_n)$  from part (i),  $h(y_n|y_n, \sigma, \alpha) = 0$ ,  $h'(x_n|y_n, \sigma, \alpha) < 0$  for all  $x_n < x_n^*$ , and  $h'(x_n|y_n, \sigma, \alpha) > 0$  for all  $x_n > x_n^*$ , it follows that  $\bar{x}(\sigma) > x_n^*$ , implying

$h(x_n|y_n, \sigma, \alpha) < 0$  for  $y_n < x_n < \bar{x}_n$  and  $h(x_n|y_n, \sigma, \alpha) > 0$  for  $x_n > \bar{x}_n$ . Recalling from part (i) that  $\Delta^{**}(x_n, y_n)$  is concave in  $x_n$  if and only if  $x_n^\alpha(1 + \alpha) \geq (\alpha - 1)(y_n^\alpha + \sigma^\alpha)$ , we can rearrange this inequality to see that it binds at  $\hat{x}(\sigma)$ . By inspection, we can then see that  $x_n < \hat{x}(\sigma)$  implies  $x_n^\alpha(1 + \alpha) < (\alpha - 1)(y_n^\alpha + \sigma^\alpha)$  and  $x_n > \hat{x}(\sigma)$  implies  $x_n^\alpha(1 + \alpha) > (\alpha - 1)(y_n^\alpha + \sigma^\alpha)$ , implying the desired result. Expressing  $\tilde{h}(\bar{x}, \sigma, y) \equiv h(\bar{x}(\sigma)|y_n, \sigma, \alpha) = 2(\bar{x} - y_n) + \sigma^\alpha(\bar{x}^{1-\alpha} - y_n^{1-\alpha}) = 0$ , we see  $\frac{\partial \tilde{h}(\bar{x}, \sigma, y)}{\partial \bar{x}} = 2 - (\alpha - 1)\bar{x}^{-\alpha}\sigma^\alpha$ ,  $\frac{\partial \tilde{h}(\bar{x}, \sigma, y)}{\partial y_n} = -2 + (\alpha - 1)y_n^{-\alpha}\sigma^\alpha$ , and  $\frac{\partial \tilde{h}(\bar{x}, \sigma, y)}{\partial \sigma} = \alpha\sigma^{\alpha-1}\left(\frac{1}{\bar{x}^{\alpha-1}} - \frac{1}{y_n^{\alpha-1}}\right) < 0$ . Next, observe  $\frac{\partial \tilde{h}(\bar{x}, \sigma, y)}{\partial \bar{x}} = 2 + (1 - \alpha)\sigma^\alpha\bar{x}^{-\alpha} > 0$ . Together, from the implicit function theorem, these inequalities imply  $\bar{x}(\sigma)$  is increasing in  $\sigma$ . By inspection, we can also readily verify that  $\hat{x}(\sigma)$  is increasing in  $\sigma$  since, holding  $\alpha > 1$  fixed,  $\hat{x}(\sigma)$  is clearly increasing in  $(y_n^\alpha + \sigma^\alpha)$  and  $(y_n^\alpha + \sigma^\alpha)$  is clearly increasing in  $\sigma$ . ■

To help convey key features of  $\Delta^{**}(x_n, y_n)$ , Proposition 16 effectively fixes the smaller attribute value, taken here to be  $y_n$ , while allowing the larger attribute value  $x_n$  to vary. Of particular relevance, if  $\sigma$  is sufficiently small in relation to  $y_n$ ,  $\Delta^{**}(x_n, y_n)$  will exhibit diminishing sensitivity and strict concavity (in  $x_n$ ) for all  $x_n \geq y_n$ . If  $\sigma$  is large in relation to  $y_n$ , however,  $\Delta^{**}(x_n, y_n)$  will instead exhibit increasing sensitivity and convexity for values of  $x_n$  that are sufficiently close to  $y_n$ .

Figure 5. The Effects of  $\sigma$  and  $\alpha$  on Contrast.



On left: increasing  $\sigma$  leads to the emergence and then the expansion of a convex region with a corresponding right-ward shift of the point at which contrast is maximally responsive to changes in  $x_n$  (for fixed  $\alpha > 1$ ). On right: with  $\alpha \leq 1$ , the contrast function will not be S-shaped, while its responsiveness becomes more concentrated over a smaller range for larger  $\alpha$  (for fixed  $\sigma > 0$ ).

Therefore  $\sigma$  determines where the direct contrast function is maximally responsive to a change in  $x_n$  relative to  $y_n$ . Since  $\Delta^{**}(x_n, y_n)$  is most responsive to changes in  $x_n$  at the threshold  $\hat{x}(\sigma)$ , the effect of increasing  $\sigma$  can also be understood here as shifting this point

of maximum responsiveness further to the right (Figure 5, left).<sup>26</sup> As noted by Rayo and Becker (2007), a bounded value function with such properties is optimal when agents are limited in their ability to discriminate small differences. While  $\sigma$  has been typically treated as a constant in the neuroscience literature (e.g. Shevell, 1977; Heeger, 1992; Louie et al., 2011), recent work suggests  $\sigma$  may arise dynamically in neural systems from the history of stimuli (LoFaro et al., 2014; Louie et al., 2014; Tymula and Glimcher, 2016; Khaw et al., 2017), thus acting as a dynamic reference point.

The constant  $\alpha > 1$  determines the extent to which the responsiveness of  $\Delta^{**}(x_n, y_n)$  is concentrated over a small range of  $x_n$ , as opposed to being dispersed over a large range. That is, as  $\alpha > 1$  increases,  $\Delta^{**}(x_n, y_n)$  becomes more responsive to changes in  $x_n$  near  $\hat{x}(\sigma)$ , but becomes less responsive for  $x_n$  further from  $\hat{x}(\sigma)$ . For example, in the limit as  $\alpha \rightarrow \infty$ ,  $\Delta^{**}(x_n, y_n)$  assumes the shape of a step-function that is infinitely responsive at  $\hat{x}(\sigma)$  but unresponsive to changes in  $x_n$  everywhere else (Figure 5, right).

## D Classifying Other Theories' Predictions

This appendix explains how other models' predictions were classified in Table 1. For each of the comparable models listed, we will describe the value function  $V(\mathbf{x}; X)$  used to classify the model's predictions and demonstrate that it generates the corresponding predictions listed in Table 1 (these value functions were also used to generate the corresponding graphs shown in Figures 2 and 4).

For clarity and to facilitate consistent comparisons across models, certain restrictions were applied to some models. For instance, we only considered deterministic versions of each model and presumed that attributes are ex-ante symmetric, so that any attribute-specific parameters or functions were taken to be the same across dimensions. These and other model-specific restrictions (discussed below) may lead to a classification of 'Y' (robustly captures the behavior) or 'N' (predicts no effect or the opposite effect) in Table 1 when a more general version of the model would imply 'S' (captures the behavior in some cases, but predicts the opposite in other cases). However, these restrictions can never prevent a 'Y' or 'N' classification (thus, if we re-created Table 1 using more general versions of each model, each classification would either remain the same or change to 'S').

The rules used to classify each prediction are then based on whether or not  $V(\mathbf{x}; X)$  as given for that model predicts the corresponding result as formalized in this paper:

- *Compromise Effect*: part (i) of Proposition 2.
- *Dominance Effect, Weak*: part (ii) of Proposition 2 with  $z_2 = x_2$ .

---

<sup>26</sup> This interpretation of  $\hat{x}(\sigma)$  follows because  $\frac{\partial \Delta^{**}(x_n, y_n)}{\partial x_n} > 0$  is increasing where  $\Delta^{**}(x_n, y_n)$  is convex and decreasing where  $\Delta^{**}(x_n, y_n)$  is concave.

- *Dominance Effect, Strict*: part (ii) of Proposition 2 with  $z_2 < x_2$ .
- *Decoy-Range Effect*: part (ii) of Proposition 3.
- *Relative Difference Effect*: Proposition 1.
- *Majority-Rule Preference Cycles*: Proposition 4.
- *Splitting Bias*: Proposition 5.
- *Alignability Effect*: Proposition 6.
- *Diversification Bias*: Proposition 7.
- *Feature Bias*: Proposition 8.

## D.1 Tversky and Simonson (1993)

For Tversky and Simonson's (1993) model, we use the following value function (for consistency, we will express other models using the notation of the basic PN model, except where new notation is needed):<sup>27</sup>

$$V(\mathbf{x}; X) = \sum_{n=1}^N x_n + \theta \cdot \sum_{\mathbf{y} \in X \setminus \mathbf{x}} \frac{\sum_n \max\{x_n - y_n, 0\}}{\sum_n \max\{x_n - y_n, y_n - x_n\}}, \quad \theta > 0. \quad (7)$$

*Compromise Effect (Y)*. Applied to (7), Assumption BCI holds if and only if  $x_1 + x_2 = y_1 + y_2$ . Taking  $x_1 + x_2 = y_1 + y_2 = 1$  (without loss of generality), we know  $z_1 + z_2 = 1 - \omega$  for some  $\omega \in (0, 1)$  since  $\mathbf{z}$  is not preferred to  $\mathbf{x}$  and  $\mathbf{y}$ . In turn, if  $\mathbf{z}$  makes  $\mathbf{x}$  a compromise, it is readily verifiable that  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  if and only if  $\frac{z_1 - x_1 + \omega}{2(z_1 - x_1) + \omega} > \frac{z_1 - y_1 + \omega}{2(z_1 - y_1) + \omega}$ , which must hold since  $z_1 > x_1 > y_1$ .

*Dominance Effect, Weak (Y) and Strict (Y)*. If  $\mathbf{x}$  asymmetrically dominates  $\mathbf{z}$ ,  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$  since  $\frac{x_1 - z_1 + x_2 - z_2}{x_1 - z_1 + x_2 - z_2} = 1 > \frac{y_2 - z_2}{y_2 - z_2 + z_1 - y_1}$  with  $z_1 > y_1$ .

*Decoy-Range Effect (N)*. It is also verifiable that the decoy-range effect is captured if  $\frac{x_1 - z_1 + x_2 - z_2}{x_1 - z_1 + x_2 - z_2} - \frac{y_2 - z_2}{y_2 - z_2 + z_1 - y_1} = 1 - \frac{y_2 - z_2}{y_2 - z_2 + z_1 - y_1} > 0$  is increasing in  $z_2$ . However, this expression is decreasing in  $z_2$  since  $y_2 > z_2$ .

*Relative Difference Effect (N)*. Using (7) and with  $\mathbf{x}'$  and  $\mathbf{y}'$  as defined in Proposition 1,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\})$  implies  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) = V(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\})$ . Thus, the relative difference effect is not captured.

*Majority-Rule Preference Cycles (N)*. Using (7),  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}\}) > V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}\})$  if and only if  $\sum_{n=1}^N x_n > \sum_{n=1}^N y_n$ . Since  $\sum_{n=1}^N x_n > \sum_{n=1}^N y_n > \sum_{n=1}^N z_n > \sum_{n=1}^N x_n$  is a contradiction, intransitive preferences (majority-rule or otherwise) are not possible.

<sup>27</sup>For simplicity, here we take  $\delta_i(t) = t$  (in their notation, see page 1885) and omit any influence of "background context" (besides the choice set) on preferences.

*Splitting Bias* (N). Using the notation in Proposition 5,  $\mathbf{x}'$  is preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}'\}$  (capturing the splitting bias) under (7) if and only if  $\frac{(x_{1a}-y_{1a})+(x_{1b}-y_{1b})-(y_2-x_2)}{x_{1a}-y_{1a}+x_{1b}-y_{1b}+y_2-x_2} > \frac{(x_1-y_1)-(y_2-x_1)}{x_1-y_1+y_2-x_2}$ , but these expressions are equal given  $x_{1a} + x_{1b} = x_1$  and  $y_{1a} + y_{1b} = y_1$ .

*Alignability Effect* (N). Using the notation in Proposition 6,  $\mathbf{x}'$  is preferred to  $\mathbf{y}'$  given  $X = \{\mathbf{x}', \mathbf{y}'\}$  (implying the alignability effect is captured) under (7) if and only if  $\frac{(x_1-y_1)+(x_2-0)-(y_2-0)}{(x_1-y_1)+(x_2-0)+(y_2-0)} > 0$ . However, under Assumption BCI,  $x_1 + x_2 = y_1 + y_2$ , implying the left-side expression is equal to zero.

*Diversification Bias* (N).  $V(\mathbf{x}; X) - V(\mathbf{x}'; X) \propto \sum_{n \leq N} \max\{\frac{A}{N} - x'_n, 0\} - \sum_{n \leq N} \max\{x'_n - \frac{A}{N}, 0\} = 0$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ ,  $\sum_{n \leq N} x'_n = A$ , and  $x_n = \frac{A}{N}$  for  $n \leq N$ . Thus, the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{x}'$ , implying the diversification bias is not captured.

*Feature Bias* (N). With  $\mathbf{x}$  and  $\mathbf{x}'$  as defined in Proposition 8,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = \frac{q}{2q} + \sum_{n \notin \{n', N\}} x_n$  under (7), implying indifference between the product  $\mathbf{x}$  with the new feature and the other product  $\mathbf{x}'$ .

## D.2 Kivetz et al. (2004a)

For Kivetz et al.'s (2004a) model, we use:

$$V(\mathbf{x}; X) = \sum_{n=1}^N (x_n - \min_{\mathbf{x}' \in X} \{x'_n\})^c, \quad 0 < c < 1. \quad (8)$$

*Compromise Effect* (Y). Noting Assumption BCI holds if and only if  $x_1 + x_2 = y_1 + y_2$  under (8), the compromise effect is captured since  $(x_1 - y_1)^c + (x_2 - z_2)^c - (y_2 - z_2)^c > 0$  must hold given  $(x_1 - y_1) + (x_2 - z_2) - (y_2 - z_2) = x_1 + x_2 - y_1 - y_2 = 0$  and  $0 < c < 1$ .

*Dominance Effect, Weak* (N) *and Strict* (Y). The dominance effect likewise holds as a result of  $(x_1 - y_1)^c + (x_2 - z_2)^c - (y_2 - z_2)^c > 0$ , provided  $z_2 < x_2$ . If  $z_2 = x_2$ , however,  $(x_1 - y_1)^c + (x_2 - z_2)^c - (y_2 - z_2)^c = (x_1 - y_1)^c - (y_2 - x_2)^c = 0$ .

*Decoy-range Effect* (Y). The decoy-range effect is captured if  $\frac{\partial}{\partial z_2} [(x_2 - z_2)^c - (y_2 - z_2)^c] = c((x_2 - z_2)^{c-1} - (y_2 - z_2)^{c-1}) > 0$ , which must hold since  $y_2 > x_2 > z_2$  and  $c < 1$ .

*Relative Difference Effect* (N). Under (8), the relative difference effect holds if  $((x_1 + k) - (y_1 + k))^c > (x_1 - y_1)^c$  for  $k > 0$ . However, both expressions are clearly equal.

*Majority-Rule Preference Cycles* (Y). We first provide an example that shows majority-rule preference cycles are possible under (8). Namely, if  $\mathbf{x} = (3, 2, 1)$ ,  $\mathbf{x}' = (2, 1, 3)$ ,  $\mathbf{x}'' = (1, 3, 2)$ , and  $c = \frac{1}{2}$ , we can then see from  $\sqrt{3-2} + \sqrt{2-1} - \sqrt{3-1} = 2 - \sqrt{2} > 0$  that, in any binary choice, the alternative that is superior on two out of three dimensions is strictly preferred. To show that minority-rule preference cycles are not possible, suppose  $\mathbf{x}$ ,  $\mathbf{x}'$ ,  $\mathbf{x}''$  satisfying the cyclical majority-dominance property where, without loss of generality,

$x_1 + x_2 + x_3 = \min_{y \in \{\mathbf{x}, \mathbf{x}', \mathbf{x}''\}} \{y_1 + y_2 + y_3\}$ ,  $\mathbf{x}'$  is superior to  $\mathbf{x}$  on two out of three attribute dimensions, and  $x_1 > x'_1$ . Hence, if a minority-rule preference cycle exists among  $\mathbf{x}$ ,  $\mathbf{x}'$ , and  $\mathbf{x}''$ , we must have  $\delta_1^c - \delta_2^c - \delta_3^c > 0$  for  $\delta_n = |x_n - x'_n|$ . Since  $x_1 + x_2 + x_3 = \min_{y \in \{\mathbf{x}, \mathbf{x}', \mathbf{x}''\}} \{y_1 + y_2 + y_3\}$ ,  $\delta_1^c - \delta_2^c - \delta_3^c \leq (\delta_2 + \delta_3)^c - \delta_2^c - \delta_3^c$ , while  $(\delta_2 + \delta_3)^c - \delta_2^c - \delta_3^c < 0$  given  $0 < c < 1$ , a minority-rule preference cycle is impossible under (8).

*Splitting Bias* (Y). The splitting bias is likewise captured under (8) since  $(x_{1a} - y_{1a})^c + (x_{1b} - y_{1b})^c > (x_{1a} + x_{1b} - y_{1a} - y_{1b})^c = (x_1 - y_1)^c$  with  $0 < c < 1$ .

*Alignability Effect* (N). The alignability effect is likewise captured under (8) since  $(y_2 - x_2)^c + x_2^c > y_2^c$  with  $0 < c < 1$ , where (as is readily verifiable) this condition ensures  $V(\mathbf{x}'; \{\mathbf{x}', \mathbf{y}'\}) > V(\mathbf{y}'; \{\mathbf{x}', \mathbf{y}'\})$  with  $\mathbf{x}'$  and  $\mathbf{y}'$  as defined in Proposition 6.

*Diversification Bias* (N). To show that (8) does not predict the diversification bias, it suffices to show an example for which the balanced allocation  $\mathbf{x}$  is not strictly preferred to some  $\mathbf{x}' \neq \mathbf{x}$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ . As one example, take  $x'_2 = \frac{2A}{N}$ ,  $x'_1 = 0$ , and  $x'_n = \frac{A}{N}$  for all  $n > 2$ . Under (8), we then have  $V(\mathbf{x}; X) = V(\mathbf{x}'; X) = \frac{A^c}{N^c}$ , implying indifference between  $\mathbf{x}$  and  $\mathbf{x}'$ .

*Feature Bias* (N). Under (8) and with  $\mathbf{x}$  and  $\mathbf{x}'$  as defined in Proposition 8,  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = q^c$ , implying there is no bias in favor of the product  $\mathbf{x}$  with the extra feature over the product  $\mathbf{x}'$  with the improvement to an existing feature.

### D.3 Bordalo et al. (2013)

For Bordalo et al.'s (2013) model, we use:

$$V(\mathbf{x}; X) = \frac{\sum_{n=1}^N \delta^{\sum_{m \neq n} \mathbb{I}[\rho_m(\mathbf{x}; X) \geq \rho_n(\mathbf{x}; X)]} \cdot x_n}{\sum_{n=1}^N \delta^{\sum_{m \neq n} \mathbb{I}[\rho_m(\mathbf{x}; X) \geq \rho_n(\mathbf{x}; X)]}}, \quad \rho_n(\mathbf{x}; X) = \left| \frac{x_n - \mu_n^X}{x_n + \mu_n^X} \right|, \quad 0 < \delta < 1, \quad (9)$$

where  $\mu_n^X \equiv ||X||^{-1} \sum_{\mathbf{x} \in X} x_n$  is the mean attribute value in  $X$  on dimension  $n$ . This formulation uses the degree-zero homogeneous salience function given in equation (4) of Bordalo et al. (2013), which is denoted here as  $\rho_n(\mathbf{x}; X)$ . For the special case of binary choice (taking  $X = \{\mathbf{x}, \mathbf{y}\}$ ) with two attributes, the model in (9) reduces to:

$$V(\mathbf{x}; X) = \begin{cases} \frac{\delta x_1 + x_2}{1 + \delta}, & \rho_1(\mathbf{x}; X) < \rho_2(\mathbf{x}; X), \\ x_1 + x_2, & \rho_1(\mathbf{x}; X) = \rho_2(\mathbf{x}; X), \\ \frac{x_1 + \delta x_2}{1 + \delta}, & \rho_1(\mathbf{x}; X) > \rho_2(\mathbf{x}; X), \end{cases} \quad \rho_n(\mathbf{x}; X) = \left| \frac{x_n - y_n}{3x_n + y_n} \right|. \quad (10)$$

As mentioned in the footnote of Table 2, Bordalo et al.'s model could be evaluated using a version in which one attribute is the price of the alternative or using a version in which all attributes represent different quality dimensions. To facilitate consistent comparisons



across models, here we consider the latter version.<sup>28</sup>

*Compromise Effect* (S). To show that (9) sometimes predicts the compromise effect and sometimes predicts the opposite, it suffices to use examples. For instance, with  $\delta = .5$ , the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{x}$  in trinary choice if  $\mathbf{x} = (3, .5)$ ,  $\mathbf{y} = (2, 1)$ , and  $\mathbf{z} = (3.2, 0)$ , in which case a compromise effect is predicted, while the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{y}$  in trinary choice if  $\mathbf{x} = (1, 2)$ ,  $\mathbf{y} = (.5, 3)$ , and  $\mathbf{z} = (1.2, 0)$ , in which case the opposite effect is predicted.

*Dominance Effect, Weak* (S) *and Strict* (S). Maintaining  $\delta = .5$ , it is similarly verifiable that the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{x}$  in trinary choice if  $\mathbf{x} = (3, .5)$ ,  $\mathbf{y} = (2, 1)$ , and  $\mathbf{z} = (2.8, 0)$ , in which case a dominance effect with a strictly dominated decoy is predicted, while the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{y}$  in trinary choice if  $\mathbf{x} = (1, 2)$ ,  $\mathbf{y} = (.5, 3)$ , and  $\mathbf{z} = (.8, 0)$ , in which case the opposite effect is predicted. In turn, the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{x}$  in trinary choice if  $\mathbf{x} = (2, 1)$ ,  $\mathbf{y} = (.5, 3)$ , and  $\mathbf{z} = (.75, 1)$ , in which case a dominance effect with a weakly dominated decoy is predicted, while the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice but prefers  $\mathbf{y}$  in trinary choice if  $\mathbf{x} = (3, .5)$ ,  $\mathbf{y} = (1, 2)$ , and  $\mathbf{z} = (1.5, .5)$ , in which case the opposite effect is predicted.

*Decoy-Range Effect* (S). Take  $\mathbf{x} = (5, 1)$ ,  $\mathbf{y} = (3, 2)$ ,  $\mathbf{z} = (4.6, 1)$ , and  $\mathbf{z}' = (4.6, .5)$  with  $\delta = .5$ . We can then compute  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{7}{3}$  and  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{11}{3}$ , implying (9) predicts the decoy-range effect in this scenario. Now take  $\mathbf{x} = (4, 3)$ ,  $\mathbf{y} = (1, 9)$ ,  $\mathbf{z} = (1.25, 3)$ , and  $\mathbf{z}' = (1, 25, 0)$ , with  $\delta = .5$ . We can then compute  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) = V(\mathbf{x}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{11}{3}$  and  $V(\mathbf{y}; \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}) = \frac{19}{3}$ , implying (9) predicts the opposite of the decoy-range effect in this scenario.

*Relative Difference Effect* (S). Maintaining  $\delta = .5$  while taking  $\mathbf{x} = (2.5, .75)$  and  $\mathbf{y} = (2, 1)$ , so that  $\mathbf{x}' = (2.5, .75 + k)$ , and  $\mathbf{y}' = (2, 1 + k)$ , we can see that the DM is indifferent in a binary choice between  $\mathbf{x}$  and  $\mathbf{y}$ . In particular, we can use (10) to compute  $\rho_1(\mathbf{x}; X) = \frac{1}{19} < \frac{1}{13} = \rho_2(\mathbf{x}; X)$  and  $\rho_1(\mathbf{y}; X) = \frac{1}{17} < \frac{1}{15} = \rho_2(\mathbf{y}; X)$ , implying  $V(\mathbf{x}; X) = \frac{.5 \cdot 2.5 + .75}{1.5} = \frac{4}{3} = \frac{.5 \cdot 2 + 1}{1.5} = V(\mathbf{y}; X)$ . Next, we can verify that in a binary choice

<sup>28</sup> Following very similar arguments and examples as those used here, it is readily verifiable that all of the Table 1 classifications for Bordalo et al.'s model would be the same for the version of their model with price as an attribute, with the possible exception of the diversification bias, which would (depending on how a model with price as an attribute was translated to the formal setting considered in Proposition 7) either: (a) no longer be testable, since allocating an equal share of an asset  $A$  to a price dimension — formally, allocating more to this dimension would mean a higher price paid — would be unnatural and in violation of the “equal returns” assumption (i.e. there would be a negative return on this dimension and a positive return on others), or (b) would be unchanged if we presume that both allocations have the same price (which may naturally be the case if the asset represents a consumption budget or a monthly contribution to a savings plan, as examples) and where the asset itself can only be allocated to the remaining quality dimensions. In this case, the salience of each alternative's price would be zero, according to  $\rho$  as defined in (9), so that the salience rankings of the quality dimensions for each alternative would be the same as the rankings with price omitted from the model.

between  $\mathbf{x}'$  and  $\mathbf{y}'$ ,  $\mathbf{y}'$  is preferred if  $k = .25$  since (10) implies  $\rho_1(\mathbf{x}'; X') = \frac{1}{19} < \frac{1}{17} = \rho_2(\mathbf{x}'; X')$  and  $\rho_1(\mathbf{y}'; X') = \frac{1}{17} > \frac{1}{19} = \rho_2(\mathbf{y}'; X')$ , implying  $V(\mathbf{x}'; X') = \frac{.5 \cdot 2.5 + .75}{1.5} = \frac{4}{3} < \frac{5}{3} = \frac{2 + .5 \cdot 1}{1.5} = V(\mathbf{y}'; X')$  with  $k = .25$ . If  $k = .5$ , however,  $\mathbf{x}'$  is preferred since (10) implies  $\rho_1(\mathbf{x}'; X') = \frac{1}{19} > \frac{1}{21} = \rho_2(\mathbf{x}'; X')$  and  $\rho_1(\mathbf{y}'; X') = \frac{1}{17} > \frac{1}{23} = \rho_2(\mathbf{y}'; X')$ , implying  $V(\mathbf{x}'; X') = \frac{2.5 + .5 \cdot .75}{1.5} = \frac{23}{12} > \frac{5}{3} = \frac{2 + .5 \cdot 1}{1.5} = V(\mathbf{y}'; X')$  with  $k = .5$ . Thus, the model captures the relative difference effect with  $k = .5$  as well as its opposite with  $k = .25$ .

*Majority-Rule Preference Cycles* (S). Take  $\mathbf{x} = (2, 1, 0)$ ,  $\mathbf{x}' = (1, 0, 2)$ ,  $\mathbf{x}'' = (0, 2, 1)$  and  $\delta = .5$ . Then a minority-rule preference cycle will exist where  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) = V(\mathbf{x}''; \{\mathbf{x}, \mathbf{x}''\}) = \frac{\delta + 2\delta^2}{1 + \delta + \delta^2} = \frac{4}{7}$  and  $V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\}) = V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}''\}) = \frac{2\delta + \delta^2}{1 + \delta + \delta^2} = \frac{5}{7} > \frac{4}{7}$ . If we instead take  $\mathbf{x} = (20, 4, 1)$ ,  $\mathbf{x}' = (4, 1, 20)$ , and  $\mathbf{x}'' = (1, 20, 4)$  while maintaining  $\delta = .5$ , then  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}'; \{\mathbf{x}', \mathbf{x}''\}) = V(\mathbf{x}''; \{\mathbf{x}, \mathbf{x}''\}) = \frac{1 + 20\delta + 4\delta^2}{1 + \delta + \delta^2} = \frac{48}{7}$  and  $V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}''; \{\mathbf{x}', \mathbf{x}''\}) = V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}''\}) = \frac{4 + \delta + 20\delta^2}{1 + \delta + \delta^2} = \frac{38}{7} < \frac{48}{7}$ , in which case a majority-rule preference cycle will exist.

*Splitting Bias* (S). Maintaining  $\delta = .5$ , take  $\mathbf{x} = (3, .5)$  and  $\mathbf{y} = (2, 1)$ , implying  $\mathbf{x}' = (x_{1a}, 3 - x_{1a}, .5)$  and  $\mathbf{y}' = (y_{1a}, 2 - y_{1a}, 1)$ . Then it is readily verifiable that the DM is indifferent in a binary choice between  $\mathbf{x}$  and  $\mathbf{y}$  under (9) and that, in a binary choice between  $\mathbf{x}'$  and  $\mathbf{y}'$ ,  $\mathbf{x}'$  is preferred if  $x_{1a} = .5$ , which captures the splitting effect, while  $\mathbf{y}'$  is preferred if  $k = .25$ , which captures the opposite.

*Alignability Effect* (S). Maintaining  $\delta = .5$ , we can verify that the alignability effect is captured under (9) if  $\mathbf{x} = (6, 1)$  and  $\mathbf{y} = (3, 2)$  but the opposite effect is predicted if  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$ .

*Diversification Bias* (S). Here, we show that the diversification bias is always captured for  $N = 2$  and never captured for  $N > 2$ . Given  $X = \{\mathbf{x}, \mathbf{x}'\}$ , with  $N = 2$  it is readily verifiable that  $V(\mathbf{x}; X) = \frac{A}{2}$ . Without loss of generality, suppose  $x'_1 < x'_2$ , implying  $x'_2 = A - x'_1$ . Then  $\rho_1(\mathbf{x}'; X) = \frac{A - 2x'_1}{A + 6x'_1}$  and  $\rho_2(\mathbf{x}'; X) = \frac{A - 2x'_1}{7A - 6x'_1}$ , implying  $\rho_1(\mathbf{x}'; X) > \rho_2(\mathbf{x}'; X)$  since  $7A - 6x'_1 > A - 2x'_1$  given  $x'_1 < x'_2 = A - x'_1$ . Hence,  $V(\mathbf{x}; X) = \frac{x'_1 + \delta(A - x'_1)}{1 + \delta} < \frac{A}{2} = V(\mathbf{x}; X)$ , as we can verify through cross-multiplication given  $2x'_1 < A$ , implying the diversification bias is captured. For  $N > 2$ , suppose  $x'_1 = x'_2 = \frac{3A}{4N}$ ,  $x'_3 = \frac{3A}{2N}$ , and  $x'_n = x_n$  for all  $n > 3$ . We then have  $\rho_1(\mathbf{x}'; X) = \rho_2(\mathbf{x}'; X) = \frac{1}{13} < \frac{1}{11} = \rho_3(\mathbf{x}'; X)$ , implying  $V(\mathbf{x}'; X) = \frac{A}{N} \cdot \frac{3(1 + \delta^2) + 2(N - 3)\delta^{N-1}}{2 + 4\delta^2 + 2(N - 3)\delta^{N-1}} > \frac{A}{N} = V(\mathbf{x}; X)$ , with the inequality holding for all  $\delta < 1$  since  $3(1 + \delta^2) > 2 + 4\delta^2$ , contradicting the diversification bias.

*Feature Bias* (Y). To show that the feature bias is robustly captured under (9), note that  $\rho_{n'}(\mathbf{x}; X) = \frac{q}{4x_{n'} + q}$  given  $X = \{\mathbf{x}, \mathbf{x}'\}$ ,  $\rho_N(\mathbf{x}; X) = \frac{1}{3}$ ,  $\rho_{n'}(\mathbf{x}'; X) = \frac{q}{4x_{n'} + 3q}$ , and  $\rho_N(\mathbf{x}'; X) = 1$ , where  $\mathbf{x}$  and  $\mathbf{x}'$  are defined as in Proposition 8. We can also see  $\rho_n(\mathbf{x}; X) = \rho_n(\mathbf{x}'; X) = 0$  for all  $n \notin \{n', N\}$ . From this, we get  $V(\mathbf{x}'; X) = \frac{\delta(x_{n'} + q) + \delta^{N-1} \sum_{n \notin \{n', N\}} x_n}{1 + \delta + (N - 2)\delta^{N-1}}$ .

If  $\frac{q}{4x_{n'}+q} = \frac{1}{3}$ , then  $V(\mathbf{x}; X) = \frac{\delta(x_{n'}+q)+\delta^{N-1}\sum_{n \notin \{n', N\}} x_n}{2\delta+(N-2)\delta^{N-1}}$ , ensuring  $V(\mathbf{x}; X) > V(\mathbf{x}'; X)$  given  $\delta < 1$ . If  $\frac{q}{4x_{n'}+q} \neq \frac{1}{3}$ , then  $V(\mathbf{x}; X) \geq \frac{\min\{x_{n'}+\delta q, \delta x_{n'}+q\}+\delta^{N-1}\sum_{n \notin \{n', N\}} x_n}{1+\delta+(N-2)\delta^{N-1}}$ , which also ensures  $V(\mathbf{x}; X) > V(\mathbf{x}'; X)$  since  $\min\{x_{n'} + \delta q, \delta x_{n'} + q\} > \delta(x_{n'} + q)$ . Thus,  $\mathbf{x}$  must be preferred to  $\mathbf{x}'$ , capturing the feature bias.

#### D.4 Koszegi and Szeidl (2013)

For Koszegi and Szeidl's (2013) model, we use:

$$V(\mathbf{x}; X) = \sum_{n=1}^N h(\max_{x' \in X} \{x'_n\} - \min_{x' \in X} \{x'_n\}) \cdot x_n, \quad (11)$$

where  $h$  is strictly increasing.

*Compromise Effect* (N). We can readily verify that the compromise effect holds under (11) if  $h(z_1 - y_1) \cdot (x_1 - y_1) > h(y_2 - z_2) \cdot (y_2 - x_2)$ . Given  $\mathbf{z}$  is not preferred to  $\mathbf{y}$ ,  $y_2 - z_2 > z_1 - y_1$  must hold, while Assumption BCI implies  $x_1 - y_1 = y_2 - x_2$ . Thus, the above condition is violated since  $h(y_2 - z_2) > h(z_1 - y_1)$  with  $h$  increasing.

*Dominance Effect — Weak* (N) *and Strict* (N). The dominance effect holds under (11) if  $h(x_1 - y_1) \cdot (x_1 - y_1) > h(y_2 - z_2) \cdot (y_2 - x_2)$ . Since  $z_2 \leq x_2$ ,  $x_1 - y_1 = y_2 - x_2$ , and  $h$  is increasing, this condition cannot hold.

*Decoy-Range Effect* (N). Given  $y_2 > x_2 \geq z_2$ ,  $\frac{\partial}{\partial z_2} [V(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) - V(\mathbf{y}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})] = (y_2 - x_2)h'(y_2 - z_2) > 0$  under (11). Thus, if the DM is indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,  $\mathbf{y}$  must be preferred to  $\mathbf{x}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}$  with  $z'_1 = z_1$  and  $z'_2 < z_2$ , which is the opposite of the decoy-range effect.

*Relative Difference Effect* (N). The relative difference effect holds if  $h((y_2+k)-(x_2+k)) \cdot ((y_2+k)-(x_2+k))$  is decreasing in  $k \geq 0$ . Since  $h((y_2+k)-(x_2+k)) \cdot ((y_2+k)-(x_2+k)) = h(y_2 - x_2) \cdot (y_2 - x_2)$ , the expression is clearly independent of  $k$ , implying the relative difference effect is not captured.

*Majority-Rule Preference Cycles* (N). To allow majority-rule preference cycles while precluding minority-rule preference cycles, it suffices to show that  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\})$  for any  $\mathbf{x}$  and  $\mathbf{z}$  satisfying  $x_1 > z_1$ ,  $x_2 > z_2$ , and  $x_1 + x_2 + x_3 = z_1 + z_2 + z_3$  with  $N = 3$ . Noting these conditions imply  $z_3 > x_3$  and  $z_3 - x_3 = x_1 + x_2 - z_1 - z_2$  while letting  $\delta_n \equiv x_n - z_n > 0$  for  $n = 1, 2$ , under (11) we have  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\}) = (h(\delta_1) - h(\delta_1 + \delta_2)) \cdot \delta_1 + (h(\delta_2) - h(\delta_1 + \delta_2)) \cdot \delta_2$ . Thus,  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\}) > 0$  cannot hold since  $h(\delta_1 + \delta_2) > h(\delta_n)$  for  $n = 1, 2$  with  $h(\cdot)$  increasing, implying majority-rule preference cycles are not captured under (11).

*Splitting Bias* (N). Letting  $\delta^a = x_{1a} - y_{1a}$  and  $\delta^b = x_{1b} - y_{1b}$ , we can see that the splitting bias is captured under (11) if  $(h(\delta^a) - h(\delta^a + \delta^b)) \cdot \delta^a + (h(\delta^b) - h(\delta^a + \delta^b)) \cdot \delta^b > 0$ , but this condition cannot hold since  $h(\delta^a + \delta^b) > \max\{h(\delta^a), h(\delta^b)\}$  with  $h$  increasing.

*Alignability Effect* (N). Since Assumption BCI implies  $x_1 + x_2 = y_1 + y_2$  under (11),  $x_1 - y_1 = y_2 - x_2$ . We can then see that the alignability effect is captured if  $(h(y_2 - x_2) - h(y_2)) \cdot (y_2 - x_2) + (h(x_2) - h(y_2)) \cdot x_2 > 0$ , which cannot hold since  $h(y_2) > \max\{h(x_2), h(y_2 - x_2)\}$  with  $h(\cdot)$  increasing.

*Diversification Bias* (N). To show that (11) does not predict the diversification bias, suppose  $x'_1 = 0$ ,  $x'_2 = \frac{2A}{N}$ , and  $x'_n = \frac{A}{N}$  for all  $n > 2$ . Then, with  $x_n = \frac{A}{N}$  for all  $n \leq N$  and  $X = \{\mathbf{x}, \mathbf{x}'\}$ , we get  $V(\mathbf{x}; X) - V(\mathbf{x}'; X) = h(\frac{A}{N} - 0) \cdot (\frac{A}{N} - 0) - h(\frac{2A}{N} - \frac{A}{N}) \cdot (\frac{2A}{N} - \frac{A}{N}) = h(\frac{A}{N}) \cdot \frac{A}{N} - h(\frac{A}{N}) \cdot \frac{A}{N} = 0$ , implying indifference between  $\mathbf{x}$  and  $\mathbf{x}'$ .

*Feature Bias* (N). With the new feature,  $\mathbf{x}$  has an effective advantage of  $h(q) \cdot q$  on dimension  $N$ . With the improved existing feature,  $\mathbf{x}'$  has an effective advantage of  $h(x_{n'} + q - x_{n'}) \cdot (x_{n'} + q - x_{n'}) = h(q) \cdot q$  on dimension  $n'$ . Since these advantages are equal under (11),  $V(\mathbf{x}; \{\mathbf{x}, \mathbf{x}'\}) = V(\mathbf{x}'; \{\mathbf{x}, \mathbf{x}'\})$  must hold, implying the feature bias is not captured.

## D.5 Bushong et al. (2017)

For Bushong et al.'s (2017) model, we use  $V(\mathbf{x}; X)$  as given in (11), except now  $h$  is strictly decreasing and  $h(z) \cdot z$  is strictly increasing in  $z$ .

*Compromise Effect* (Y). As before, the compromise effect holds if  $h(z_1 - y_1) \cdot (x_1 - y_1) > h(y_2 - z_2) \cdot (y_2 - x_2)$ . Given  $\mathbf{z}$  is not preferred to  $\mathbf{y}$ ,  $y_2 - z_2 > z_1 - y_1$  must hold, while Assumption BCI implies  $x_1 - y_1 = y_2 - x_2$ . Thus, the above condition holds (ensuring a compromise effect) since  $h(y_2 - z_2) < h(z_1 - y_1)$  given  $h$  is decreasing.

*Dominance Effect, Weak* (N) *and Strict* (Y). The dominance effect holds if  $h(x_1 - y_1) \cdot (x_1 - y_1) > h(y_2 - z_2) \cdot (y_2 - x_2)$ . Since  $x_1 - y_1 = y_2 - x_2$  and  $h$  is decreasing, the condition holds for  $z_2 < x_2$  but not for  $z_2 = x_2$ . Thus, the dominance effect is captured for a strictly dominated decoy  $\mathbf{z}$  but not if  $\mathbf{z}$  is only weakly dominated.

*Decoy-Range Effect* (Y). As seen,  $\frac{\partial}{\partial z_2} [V(\mathbf{x}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}) - V(\mathbf{y}, \{\mathbf{x}, \mathbf{y}, \mathbf{z}\})] = (y_2 - x_2)h'(y_2 - z_2)$  given  $y_2 > x_2 \geq z_2$  under (11), except now  $(y_2 - x_2)h'(y_2 - z_2) < 0$  since  $h'(y_2 - z_2) < 0$ . Thus, with indifference between  $\mathbf{x}$  and  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,  $\mathbf{x}$  will be preferred to  $\mathbf{y}$  given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}'\}$  with  $z'_1 = z_1$  and  $z'_2 < z_2$ , capturing the decoy-range effect.

*Relative Difference Effect* (N). Same as Koszegi and Szeidl (2013) — see above.

*Majority-Rule Preference Cycles* (Y). To allow majority-rule preference cycles while precluding minority-rule preference cycles, it suffices to show  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) > V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\})$  for any  $\mathbf{x}$  and  $\mathbf{z}$  satisfying  $x_1 > z_1$ ,  $x_2 > z_2$ , and  $x_1 + x_2 + x_3 = z_1 + z_2 + z_3$  with

$N = 3$ . Noting these conditions imply  $z_3 > x_3$  and  $z_3 - x_3 = x_1 + x_2 - z_1 - z_2$  while letting  $\delta_n \equiv x_n - z_n > 0$  for  $n = 1, 2$ , under (11) we have  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\}) = (h(\delta_1) - h(\delta_1 + \delta_2)) \cdot \delta_1 + (h(\delta_2) - h(\delta_1 + \delta_2)) \cdot \delta_2$ . Thus,  $V(\mathbf{x}, \{\mathbf{x}, \mathbf{z}\}) - V(\mathbf{z}, \{\mathbf{x}, \mathbf{z}\}) > 0$  must hold since  $h(\delta_1 + \delta_2) < h(\delta_n)$  for  $n = 1, 2$  given  $h$  is decreasing, implying majority-rule preference cycles are robustly captured.

*Splitting Bias* ( $\mathbf{Y}$ ). Given  $\delta^a = x_{1a} - y_{1a}$  and  $\delta^b = x_{1b} - y_{1b}$ , the splitting bias is captured if  $(h(\delta^a) - h(\delta^a + \delta^b)) \cdot \delta^a + (h(\delta^b) - h(\delta^a + \delta^b)) \cdot \delta^b > 0$ , which must hold since  $h(\delta^a + \delta^b) < \min\{h(\delta^a), h(\delta^b)\}$  with  $h$  decreasing.

*Alignability Effect* ( $\mathbf{Y}$ ). Since Assumption BCI implies  $x_1 + x_2 = y_1 + y_2$  under (11),  $x_1 - y_1 = y_2 - x_2$ . Using this relation, we can then see that the alignability effect is captured if  $(h(y_2 - x_2) - h(y_2)) \cdot (y_2 - x_2) + (h(x_2) - h(y_2)) \cdot x_2 > 0$ , which must hold since  $h(y_2) < \min\{h(x_2), h(y_2 - x_2)\}$  with  $h$  decreasing.

*Diversification Bias* ( $\mathbf{N}$ ). Same as Koszegi and Szeidl (2013) — see above.

*Feature Bias* ( $\mathbf{N}$ ). Same as Koszegi and Szeidl (2013) — see above.

## E Model Restrictions in Figure 2

This appendix describes parametric and functional form assumptions used to create the graphs in Figure 2 for the Tversky and Simonson (1993), Kivetz et al. (2004a), and Bordalo et al. (2013) models. As noted in the text, these graphs depicted each model's predicted effect of adding a third alternative  $\mathbf{z}$  on the DM's preference between two alternatives,  $\mathbf{x}$  and  $\mathbf{y}$ , where the DM was indifferent between  $\mathbf{x}$  and  $\mathbf{y}$  in binary choice. With one exception (addressed below), we used  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$  while expressing the DM's trinary-choice preference between  $\mathbf{x}$  and  $\mathbf{y}$  in terms of  $\mathbf{z}$ 's attribute values,  $z_1$  and  $z_2$ . In turn, the parametric and functional form assumptions described below were selected due to their simplicity and adherence to the more general restrictions of the model to which they were applied.

*Tversky and Simonson (1993)*. To generate the graph for Tversky and Simonson's (1993) model, we used the value function (7) described in Appendix D. For any  $\theta > 0$ , it is then readily verifiable under (7) that, with  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$  and  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,

$$V(\mathbf{x}; X) - V(\mathbf{y}; X) \propto \frac{\max\{2 - z_1, 0\} + \max\{1 - z_2, 0\}}{|2 - z_1| + |1 - z_2|} - \frac{\max\{1 - z_1, 0\} + \max\{2 - z_2, 0\}}{|1 - z_1| + |2 - z_2|},$$

which generates the regions shown in Figure 2 for Tversky and Simonson's (1993) model.

*Kivetz et al. (2004a)*. To generate the graph for Kivetz et al.'s (2004a) model, we used the value function (8) described in Appendix D. It is then readily verifiable under (8)

that, with  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$  and  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,

$$V(\mathbf{x}; X) - V(\mathbf{y}; X) = \sum_{n=1}^2 (3 - n - \min\{1, z_n\})^c - (n - \min\{1, z_n\})^c,$$

which, taking any  $c \in (0, 1)$ , generates the regions shown in Figure 2 for the Kivetz et al. (2004a) model.

*Bordalo et al. (2013) — Two Quality Attributes.* To generate the graph for Bordalo et al.'s (2013) model with alternatives defined on two quality dimensions, we used the value function (9) described in Appendix D. It is then readily verifiable under (9) that, with  $\mathbf{x} = (2, 1)$  and  $\mathbf{y} = (1, 2)$  and  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ ,

$$V(\mathbf{x}; X) - V(\mathbf{y}; X) = \frac{2\delta^{g_x(\mathbf{z})} + \delta^{1-g_x(\mathbf{z})}}{\delta^{g_x(\mathbf{z})} + \delta^{1-g_x(\mathbf{z})}} - \frac{\delta^{g_y(\mathbf{z})} + 2\delta^{1-g_y(\mathbf{z})}}{\delta^{g_y(\mathbf{z})} + \delta^{1-g_y(\mathbf{z})}},$$

where  $g_x(\mathbf{z}) \equiv \mathbb{I}\left[\frac{|3-z_1|}{9+z_1} > \frac{z_2}{6+z_2}\right]$  and  $g_y(\mathbf{z}) \equiv \mathbb{I}\left[\frac{z_1}{6+z_1} > \frac{|3-z_2|}{9+z_2}\right]$ . We can then see that, for any  $\delta \in (0, 1)$ , these expressions generate the regions shown in Figure 2.

*Bordalo et al. (2013) — Price and Quality.* To generate the graph for Bordalo et al.'s (2013) model with alternatives defined by its price and a single quality attribute, we used  $\mathbf{x} = (p_x, q_x) = (1, 1)$  and  $\mathbf{y} = (p_y, q_y) = (2, 2)$  while otherwise applying the same restrictions in (9), which can still be applied with price as an attribute simply by treating the price of  $\mathbf{z} \in \{\mathbf{x}, \mathbf{y}\}$  as a negative quality attribute with value  $-p_z$ . Given  $X = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ , the predicted value difference between  $\mathbf{x}$  and  $\mathbf{y}$  in trinary choice is then

$$V(\mathbf{x}; X) - V(\mathbf{y}; X) = \frac{\delta^{1-g_x(\mathbf{z})} - \delta^{g_x(\mathbf{z})}}{\delta^{g_x(\mathbf{z})} + \delta^{1-g_x(\mathbf{z})}} - 2 \cdot \frac{\delta^{1-g_y(\mathbf{z})} - \delta^{g_y(\mathbf{z})}}{\delta^{g_y(\mathbf{z})} + \delta^{1-g_y(\mathbf{z})}},$$

where now  $g_x(\mathbf{z}) \equiv \mathbb{I}[p_z > q_z]$  and  $g_y(\mathbf{z}) \equiv \mathbb{I}\left[\frac{|3-p_z|}{9+p_z} < \frac{|3-q_z|}{9+q_z}\right]$ . In turn, these expressions (again, with any  $\delta \in (0, 1)$ ) generate the regions shown in Figure 2.