RAN KIVETZ, ODED NETZER, and V. SRINIVASAN*

The compromise effect denotes the finding that brands gain share when they become the intermediate rather than extreme option in a choice set. Despite the robustness and importance of this phenomenon, choice modelers have neglected to incorporate the compromise effect in formal choice models and to test whether such models outperform the standard value maximization model. In this article, the authors suggest four context-dependent choice models that can conceptually capture the compromise effect. Although the models are motivated by theory from economics and behavioral decision research, they differ with respect to the particular mechanism that underlies the compromise effect (e.g., contextual concavity versus loss aversion). Using two empirical applications, the authors (1) contrast the alternative models and show that incorporating the compromise effect by modeling the local choice context leads to superior predictions and fit compared with the traditional value maximization model and a stronger (naive) model that adjusts for possible biases in utility measurement, (2) generalize the compromise effect by demonstrating that it systematically affects choice in larger sets of products and attributes than has been previously shown, (3) show the theoretical and empirical equivalence of loss aversion and local (contextual) concavity, and (4) demonstrate the superiority of models that use a single reference point over "tournament models" in which each option serves as a reference point. They discuss the theoretical and practical implications of this research as well as the ability of the proposed models to predict other behavioral context effects.

Alternative Models for Capturing the Compromise Effect

Everything in moderation, including moderation.
—Source unknown

The compromise effect, whereby brands gain share when they become intermediate options in a choice set, is among the most important and robust phenomena documented in behavioral research in marketing. First demonstrated by Simonson (1989), the compromise effect has since been investigated in many studies (e.g., Benartzi and Thaler 2002; Chernev 2004; Dhar, Nowlis, and Sherman 2000; Drolet 2002; Nowlis and Simonson 2000). Although these studies provide important insights into the effect's antecedents and moderators, the questions whether the compromise effect can be incorporated in formal choice models and whether doing so increases modelers' ability to predict consumer choice have been neglected. This issue also has important practical implications, because accounting for the compromise effect in models that predict consumer demand can enable marketers to construct strategically choice sets that increase the attractiveness and purchase likelihood of designated (high-margin) options.

In this article, we build on several theoretical mechanisms that may underlie the compromise effect, and we propose four alternative choice models that can conceptually capture the effect. We contrast these models using two empirical applications; some key findings indicate the following:

- The importance of modeling the local choice context, as implied by the superiority of the alternative models over the context-independent value-maximization model and another model that incorporates global concavity (i.e., diminishing sensitivity) and adjusts for possible biases in utility measurement;

*Ran Kivetz is Sidney Taurel Associate Professor of Business (e-mail: rk566@columbia.edu), and Oded Netzer is Assistant Professor of Marketing (e-mail: on2110@columbia.edu), Graduate School of Business, Columbia University. V. "Seenu" Srinivasan is Adams Distinguished Professor of Management, Graduate School of Business, Stanford University (e-mail: srinivasan_seenu@gsb.stanford.edu). The authors are grateful for helpful comments and suggestions received from Professor Donald R. Lehmann, participants in the Stanford Business School Marketing Seminar Series, and the anonymous *JMR* reviewers.

Journal of Marketing Research
Vol. XLI (August 2004), 237–257

- The (empirical) equivalence of loss aversion and local (contextual) concavity;
- The superiority of models that use a single reference point over "tournament models" in which each option serves as a reference point; and
- The generalization of the compromise effect to larger choice sets and dimensional spaces (a market with five product alternatives described on four attributes) than has been previously shown.

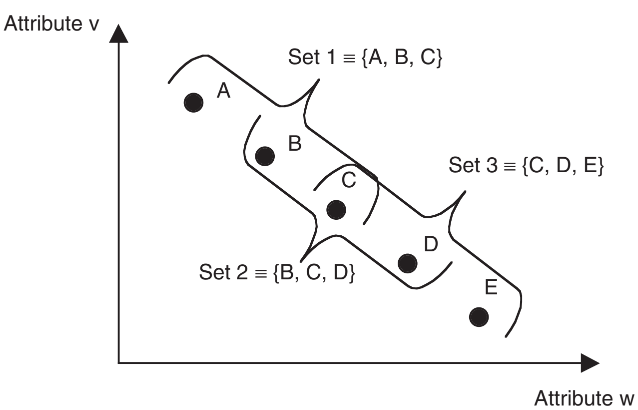
The article is organized as follows: We first review the compromise effect and its implications for consumer choice. We then employ several theoretical mechanisms that may underlie the compromise effect and propose four choice models that are designed to capture the effect. In "Empirical Application 1," we discuss the first study, its results, and the calibration and validation of the alternative models. In "Empirical Application 2," we report a second study, which generalizes the compromise effect and tests the alternative models in a choice setting that is more complex than previous demonstrations of the effect. Finally, we conclude by discussing the theoretical and practical implications of this research.

THE COMPROMISE EFFECT

The compromise effect denotes the phenomenon that the share of a product is enhanced when it is the intermediate option in a choice set and diminished when it is an extreme option (e.g., Simonson 1989; Simonson and Tversky 1992). Thus, "compromise" implies a context effect, whereby the attractiveness of an option is greater in the context of a triplet in which it is the intermediate (compromise) option than in a triplet in which it is an extreme. For example, the share of Option B relative to that of Option C (see Figure 1) is greater in the set {A, B, C} than in the set {B, C, D}; following Simonson and Tversky (1992), we denote this as $P_A(B; C) > P_D(B; C)$. Similarly, compromise implies that $P_B(C; D) > P_E(C; D)$.

The compromise effect has been documented in many studies and across a wide range of domains and product categories, such as apartments, investment portfolios, and mouthwashes (e.g., Benartzi and Thaler 2002; Dhar and Simonson 2003; Lehmann and Pan 1994; Simonson and Nowlis 2000). The effect has been found to be highly robust and of substantial magnitude. For example, Simonson (1989) reports that across five product categories, alternatives gained an average 17.5% (absolute) market share when they became the compromise option in a choice set. We find similarly strong compromise effects in the empirical applications that we report subsequently.

Figure 1
A SCHEMATIC ILLUSTRATION OF THE COMPROMISE EFFECT



Notes: Both attributes are such that preference increases with the attribute value, when the other attribute is held constant.

Implications of the Compromise Effect for Consumer Choice

The compromise effect has important theoretical implications for consumer choice and its modeling. Tversky and Simonson (1993) show that under a highly plausible condition (called the "ranking condition"), the compromise effect is inconsistent with a (possibly) heterogeneous set of value-maximizing consumers. In particular, value maximization implies a "betweenness inequality," such that the addition of an extreme option draws more share away from the intermediate (and more similar) option than from the other extreme (and less similar) option (Tversky and Simonson 1993, pp. 1180, 1188). Thus, betweenness inequality implies that $P(B; C) > P_A(B; C)$ (see Figure 1), whereas compromise predicts the opposite. Furthermore, Hutchinson, Kamakura, and Lynch (2000) show that aggregation biases are unlikely to explain between-subjects compromise effects.

The compromise effect also has significant implications for positioning, branding, and competitive strategies. For example, it suggests that the introduction of a top-of-the-line product draws disproportionately more share away from a lower-end product than from a more similar, intermediate option. Thus, marketers who wish to promote high-margin products (typically higher-priced options) may be able to do so by introducing another alternative that is even more expensive.

Furthermore, product assortments presented according to the underlying structure of the compromise effect appear to be pervasive, if not ubiquitous, in the marketplace. Anecdotal evidence suggests that the basic design of three options defined on two attributes is present in many categories, including choices among insurance plans, soft drinks, audio speakers, cellular telephone subscriptions, and so on (for examples of two sets, see Figure 2). In addition to this simple design, customers often face market choices among triplets with a trade-off between price and *several* quality-related dimensions that are highly correlated environmentally (e.g., different options such as Toyota Camry SE, LE, and XLE). Customers are likely to simplify such choices by construing the (correlated) quality dimensions as one "meta-attribute" and by making their decision on the basis of price versus overall product quality (Green and Srinivasan 1978; Wright 1975).

Customers often face larger sets of options. Nevertheless, such market choices can produce compromise effects, because they often contain sets of Pareto-optimal alternatives that require customers to make attribute trade-offs. Indeed, in a subsequent section, we demonstrate strong compromise effects in larger sets of options and attributes.

In summary, the robustness of the compromise effect and its important theoretical and practical implications emphasize the need for formal choice models that can predict the

Figure 2
EXAMPLES OF REAL-WORLD ASSORTMENTS CONSISENT WITH A COMPROMISE STRUCTURE

effect. Accordingly, regardless of whether the compromise effect can be justified normatively (for related discussions, see Prelec, Wernerfelt, and Zettelmeyer 1997; Wernerfelt 1995; cf. Drolet, Simonson, and Tversky 2000; Tversky and Simonson 1993), the standard utility model must be modified to capture it. This is the goal of this article. Next, we present four alternative models that incorporate the compromise effect by modeling the local choice context.

THEORY AND MODELS

In this section, we introduce four context-dependent multiattribute choice models that are designed to capture the compromise effect. The models are motivated by theory from economics and behavioral decision research. However, it should be noted that our main goal is to construct and test "as-if" models that improve predictive validity rather than to explore underlying decision processes.

A key characteristic of all models is that they view choice as a constructive process (e.g., Bettman, Luce, and Payne 1998; Payne, Bettman, and Johnson 1992), whereby consumers modify their preferences on the basis of the local choice set. Furthermore, in line with the suggestions of Drolet, Simonson, and Tversky (2000) and Hardie, Johnson, and Fader (1993), the four alternative models incorporate both relative (i.e., reference-dependent) and absolute (i.e., global) elements of consumer choice. In particular, our modeling approach assumes that the utilities (partworths) of attribute levels are known and have been measured at a global (context-independent) level, with methods such as conjoint analysis. The global partworth functions can assume any shape (e.g., linear, concave, convex). The models then transform the context-independent partworth utilities according to the local context (i.e., based on the relationship among options in a choice set). Thus, we subsequently illustrate the alternative models using equations and graphs that operate at the subjective utility space rather than the objective attribute space. Moreover, the alternative models consist of individual-level utility functions and therefore account for heterogeneity through the estimated context-independent partworths.

It is important to emphasize that the proposed models are in no way limited to a particular method of preference modeling (e.g., partworth function versus vector [linear] model) or preference measurement technique (e.g., full-profile versus self-explicated approach). That is, when discussing the models in this section, we use the term "partworth" in the most general sense, to denote the utility or worth of a specific attribute level for an individual consumer. However, our models assume additivity such that the overall utility for the product is the sum of the partworths for the product's levels on the different attributes (Green and Srinivasan 1978, p. 105). We also assume that the attributes are such that preference increases (or decreases) with the levels of an attribute (i.e., it is not of the ideal-point type).

The models are "general compromise" models in the sense that they can capture any form of compromise (or extremeness aversion), that is, compromise of either equal or different magnitude across attributes.[1] It should also be noted that the models are not limited in terms of the number of attributes or choice-set size, unlike the extant behavioral literature on the compromise effect, which uses at most three options and two attributes. Indeed, in our second empirical application, we test the models in the context of larger choice sets (five options described on four attributes). Next, we discuss each model and its underlying conceptual mechanism.

[1]Simonson and Tversky (1992) label the finding that intermediate options fare better than extreme options as "extremeness aversion," and they argue that it leads to two types of effects: "compromise," which represents cases in which both attributes exhibit extremeness aversion, and "polarization," which represents cases in which only one attribute exhibits such an effect.

The Contextual Concavity Model

A robust empirical generalization about human perception and decision making is that of diminishing returns, or sensitivity (e.g., Meyer and Johnson 1995). Specifically, the mapping of objective attribute values (or "gains") onto psychological value is a concave function (e.g., Thaler 1985; Tversky and Kahneman 1991). We suggest that the compromise effect can be mathematically modeled by combining the notions of concavity and context dependence, or in other words by "contextual concavity." More specifically, according to the contextual concavity model (CCM), the deterministic component of utility of alternative j (for consumer i) equals the sum across attributes of concave functions of the partworth gains between this alternative and the alternative with the minimum partworth in the (local) choice-set S:[2]

$$M_{ij}^{S} = \sum_{k}\left(P_{ijk} - P_{i,min,k}^{S}\right)^{c_k}, \tag{1}$$

where

M_{ij}^{S} = the deterministic component of utility of alternative j in context (choice-set) S for consumer i,
P_{ijk} = the partworth (utility) of the level of attribute k of alternative j for consumer i,
$P_{i,min,k}^{S}$ = the partworth of attribute k of the alternative with the lowest partworth on this attribute in choice-set S for consumer i, and
c_k = the concavity parameter of attribute k.

Note that the partworths may already be a concave function of the objective attribute values. Contextual concavity adds another layer of concavity on top of that, because the model operates at the subjective partworth (utility) space. The overall utility of alternative j in context S includes the sum of the deterministic part (i.e., M_{ij}^{S}) and an error term ε_{ij} that captures the unobserved (to the researcher) component of utility.

If we assume that the error term ε_{ij} is distributed i.i.d. of the extreme value type, the probability that consumer i will chose alternative j in the context of S follows the multinomial logit model (McFadden 1974):[3]

$$Pr_i(j|S) = \frac{\exp(bM_{ij}^{S})}{\sum_{h}\exp(bM_{ih}^{S})}, \tag{2}$$

where j, h ∈ S. The parameters for estimation in the CCM are the logit scale parameter b and the concavity parameters $\{c_k\}$. In general, the CCM requires estimation of d + 1 parameters, where d is the number of product dimensions (attributes). We expect that the parameter b will be positive, thus capturing the positive effect of utility on choice. In addition, we expect that the parameters c_k will be smaller than one, thus capturing the contextual concavity in consumers' preference structure. Furthermore, the CCM can capture situations in which the magnitude of the compromise effect varies across attributes by allowing the (smaller-than-one) concavity parameters to differ by attribute.

It is noteworthy that if $c_k = 1$ for all k, we have the simple multinomial logit based on the value maximization model (VMM). This is true because $M_{ij}^{S} = \sum_k(P_{ijk} - P_{i,min,k}^{S})^1 = \sum_k P_{ijk} - \sum_k P_{i,min,k}^{S}$, which equals the commonly used partworth additive utility function minus a constant (the last term does not depend on j). Because the multinomial logit probability is invariant to an additive constant, the proposed model (given that $c_k = 1$ for all k) and the traditional VMM are identical. If $c_k > 1$ for all k, there is "extremeness seeking" (i.e., a preference for extreme options rather than compromise ones).

Figure 3 illustrates how the CCM captures the compromise effect using the example of three portable personal computers (PCs), {A, B, C}, that vary in terms of their speed (s) and memory (m). We used this trinary choice set in our first empirical application (described in the next section). Therefore, Figure 3 employs the measured partworths of a real participant and the concavity parameters actually estimated. Panel A of Figure 3 represents the attribute speed, whereas Panel B represents memory. In Panels A and B, the x-axis depicts the three options, and the y-axis depicts the (modeled) deterministic component of utility for the relevant attribute of each option; that is, the relative partworth gain, $M_{ij}^{S}(k) = (P_{ijk} - P_{i,min,k}^{S})^{c_k}$. For simplicity of exposition, we depict the utility for consumer i who has linear partworth functions and equal importances for speed and memory (i.e., $P_{i,max,s}^{S} - P_{i,min,s}^{S} = P_{i,max,m}^{S} - P_{i,min,m}^{S}$).[4] The (45 degree) diagonal lines, which represent the VMM, are based on Equation 1 and assume concavity parameters of 1 (i.e., $c_s = c_m = 1$, for speed and memory, respectively). In contrast, the CCM (represented with concave lines) incorporates compromise with concavity parameters that are smaller than 1 (i.e., $c_s, c_m < 1$).

Panel C of Figure 3 aggregates the modeled utilities across the two attributes to form graphs of the overall deterministic component of utility of each alternative in context S (i.e., M_{ij}^{S}). It is clear that (for this specific consumer) value maximization implies A ~ B ~ C, because the sum of the relative partworth gains is equal across the three options (i.e., 0 + 11.1 = 5.55 + 5.55 = 11.1 + 0, respectively), as is indicated by the (flat) aggregate VMM graph.[5] Conversely, the CCM implies that A ≺ B ≻ C, because the sum of the concave partworth gains is highest for option B (i.e., 0 + 2.3 < 2.2 + 1.8 > 3.1 + 0, respectively).

More generally, the relative preference for an intermediate option (with respect to the extreme options) is greater in the CCM than in the VMM. The CCM operates as if the VMM partworth line is pulled downward from its highest point; all partworths (except the lowest one) are lowered, but higher partworths are lowered relatively more; thus, the relative preference for intermediate options is enhanced.

A noteworthy disadvantage of the CCM is that it distorts the relative importance weights of attributes. The *importance weight* of an attribute is defined as the range of variation of partworths for that attribute (see, e.g., Green and Srinivasan 1978). If the contextual concavity parameter of

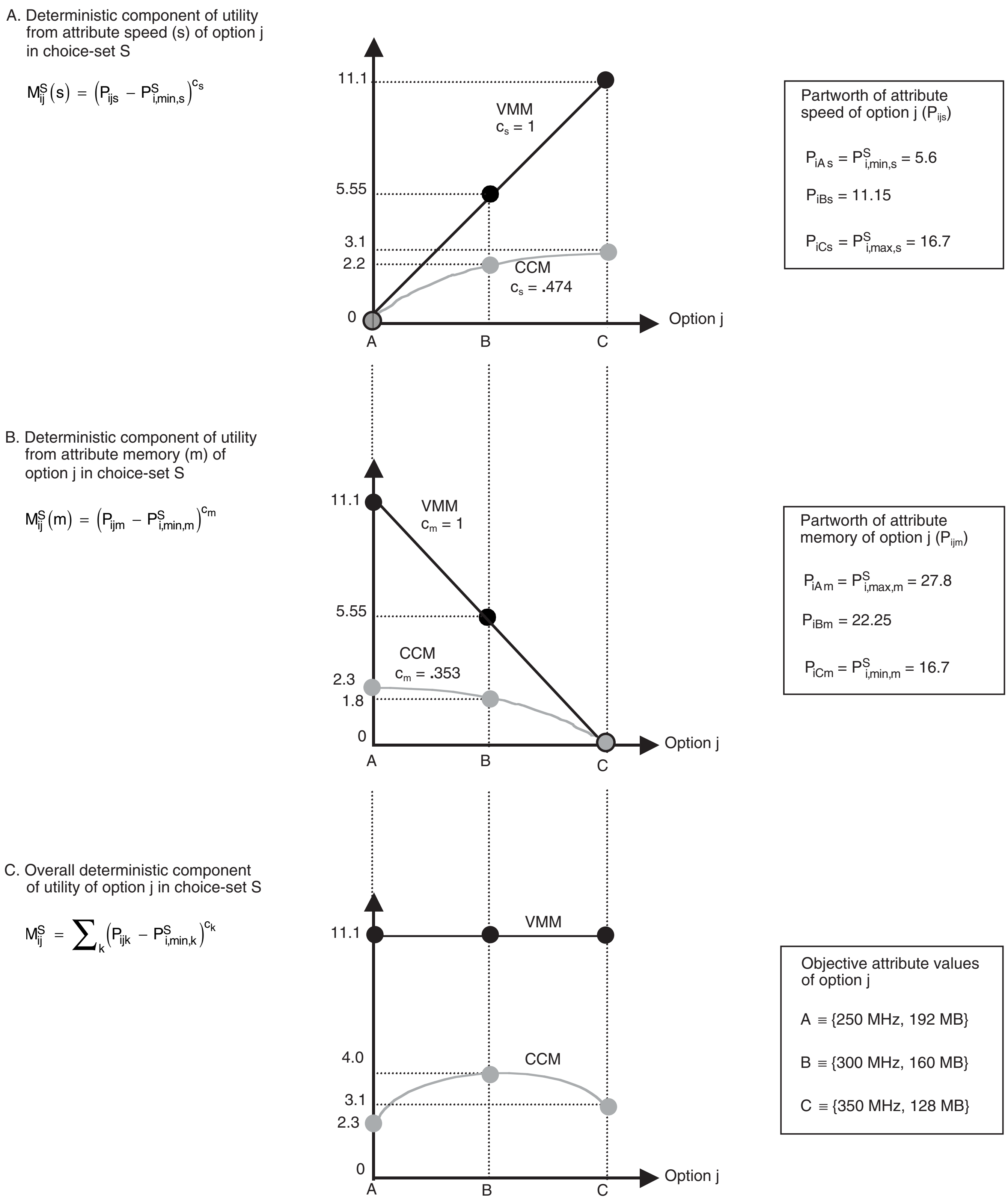
[2]We rescaled the partworth utilities to range from 0 to 100. For all practical purposes, partworth gains ($P_{ijk} - P_{i,min,k}^{S}$) are either equal to 0 or greater than 1.

[3]We assume that the error term ε_{ij} is distributed i.i.d. with the extreme value distribution in all subsequent models as well.

[4]The ensuing analysis holds even when the partworth functions are nonlinear and/or the importances are unequal.

[5]In what follows, the notation A ~ B means that $M_{iA}^{S} = M_{iB}^{S}$; similarly A ≻ [≺] B means that M_{iA}^{S} > [<] M_{iB}^{S}.

Figure 3
THE CONTEXTUAL CONCAVITY MODEL: ATTRIBUTE-SPECIFIC AND AGGREGATE UTILITY GRAPHS
(FOR A PARTICULAR CONSUMER WITH EQUAL ATTRIBUTE IMPORTANCES)



an attribute is smaller (i.e., greater concavity) than that of another attribute, then the former attribute's relative importance is diminished (see Figure 3). Consequently, if the CCM were to predict better than the VMM, a rival explanation would be that the CCM transforms the stated attribute importances (obtained by conjoint analysis) to importances revealed through choices. Next, we revise the CCM to avoid this issue.

The Normalized Contextual Concavity Model

In the normalized contextual concavity model (NCCM), we normalize the concave partworth gain of each attribute by the attribute's weight (see Equation 3). This enables us to retain the original attribute importance weights while incorporating compromise through contextual concavity.

$$\text{(3)} \qquad M_{ij}^{S} = \sum_k \left(P_{i,\max,k}^{S} - P_{i,\min,k}^{S}\right) \times \left[\left(P_{ijk} - P_{i,\min,k}^{S}\right) \Big/ \left(P_{i,\max,k}^{S} - P_{i,\min,k}^{S}\right)\right]^{c_k}.$$

As in the original CCM, the probability that consumer i will chose alternative j according to the NCCM has a multinomial logit structure (see Equation 2). The parameters for estimation and their interpretations are similar to the ones used in the CCM.

To better convey the mechanics of the NCCM, we again present attribute-specific and total utility graphs (see Figure 4) and use the example discussed previously (i.e., set {A, B, C} and the context-independent partworths of an actual consumer). The (45 degree) diagonal lines shown in Panels A and B of Figure 4 represent the VMM and are based on Equation 3 (assuming that $c_s = c_m = 1$). Conversely, the concave NCCM lines capture compromise by c_s, $c_m < 1$. As is evident in Figure 4, compared with the VMM, the NCCM yields higher attribute utilities and thus a higher total utility for all intermediate alternatives. However, for both $P_{ijk} = P_{i,\min,k}^{S}$ and $P_{ijk} = P_{i,\max,k}^{S}$, the utility from attribute k is the same according to the NCCM and the VMM. Next, we suggest an alternative model for incorporating the compromise effect based on a framework proposed by Tversky and Simonson (1993).

The Relative Advantage Model

Although previous research has neglected to examine empirically choice models that capture the compromise effect, Tversky and Simonson (1993) take a step in this direction. Specifically, they propose the following modeling framework in which the compromise effect is incorporated through a linear combination of two elements:

$$\text{(4)} \qquad M_{ij}^{S} = bV_{ij} + qR_{ij}^{S}.$$

The first element (i.e., V_{ij}) is the context-independent value of option j for consumer i. This component equals the sum of the attribute partworths (i.e., $\Sigma_k P_{ijk}$), which are independent of the local choice set. The second element (i.e., R_{ij}^{S}) captures the impact of the relative position of option j with respect to all other options in choice-set S. This component equals the sum of the relative advantages of option j with respect to each of the other options in the choice set and is defined as

$$\text{(5)} \qquad R_{ij}^{S} = \sum_{h \neq j} R_i(j, h) = \sum_{h \neq j} \frac{A_i(j, h)}{A_i(j, h) + D_i(j, h)},$$

where

$j, h \in S$;

$A_i(j, h) = \Sigma_k A_{ik}(j, h) = \Sigma_k (P_{ijk} - P_{ihk}) \times 1(P_{ijk} > P_{ihk})$, that is, the advantage of option j relative to option h for consumer i;

$1(P_{ijk} > P_{ihk})$ = an indicator function that equals 1 if $P_{ijk} > P_{ihk}$ and equals 0 otherwise; and

$D_i(j, h) = \Sigma_k D_{ik}(j, h)$, that is, the disadvantage of option j relative to option h for consumer i.

Following Tversky and Simonson (1993), we define the disadvantage of option j with respect to option h on each attribute k (i.e., $D_{ik}(j, h)$) as an increasing convex function of the corresponding advantage of h relative to j (i.e., $A_{ik}(h, j)$), such that $D_{ik}(j, h) > A_{ik}(h, j)$. Although Tversky and Simonson (1993) do not specify a particular function, a general form that satisfies these requirements is the following power function:[6]

$$\text{(6)} \qquad D_{ik}(j, h) = [A_{ik}(h, j) + L_k \times A_{ik}(h, j)^{\psi_k}] \times 1(P_{ihk} > P_{ijk})$$
$$= [(P_{ihk} - P_{ijk}) + L_k \times (P_{ihk} - P_{ijk})^{\psi_k}] \times 1(P_{ihk} > P_{ijk}).$$

The probability that consumer i will chose alternative j in the context of S has the multinomial logit form

$$\text{(7)} \qquad \Pr_i(j|S) = \frac{\exp(M_{ij}^{S})}{\sum_h \exp(M_{ih}^{S})},$$

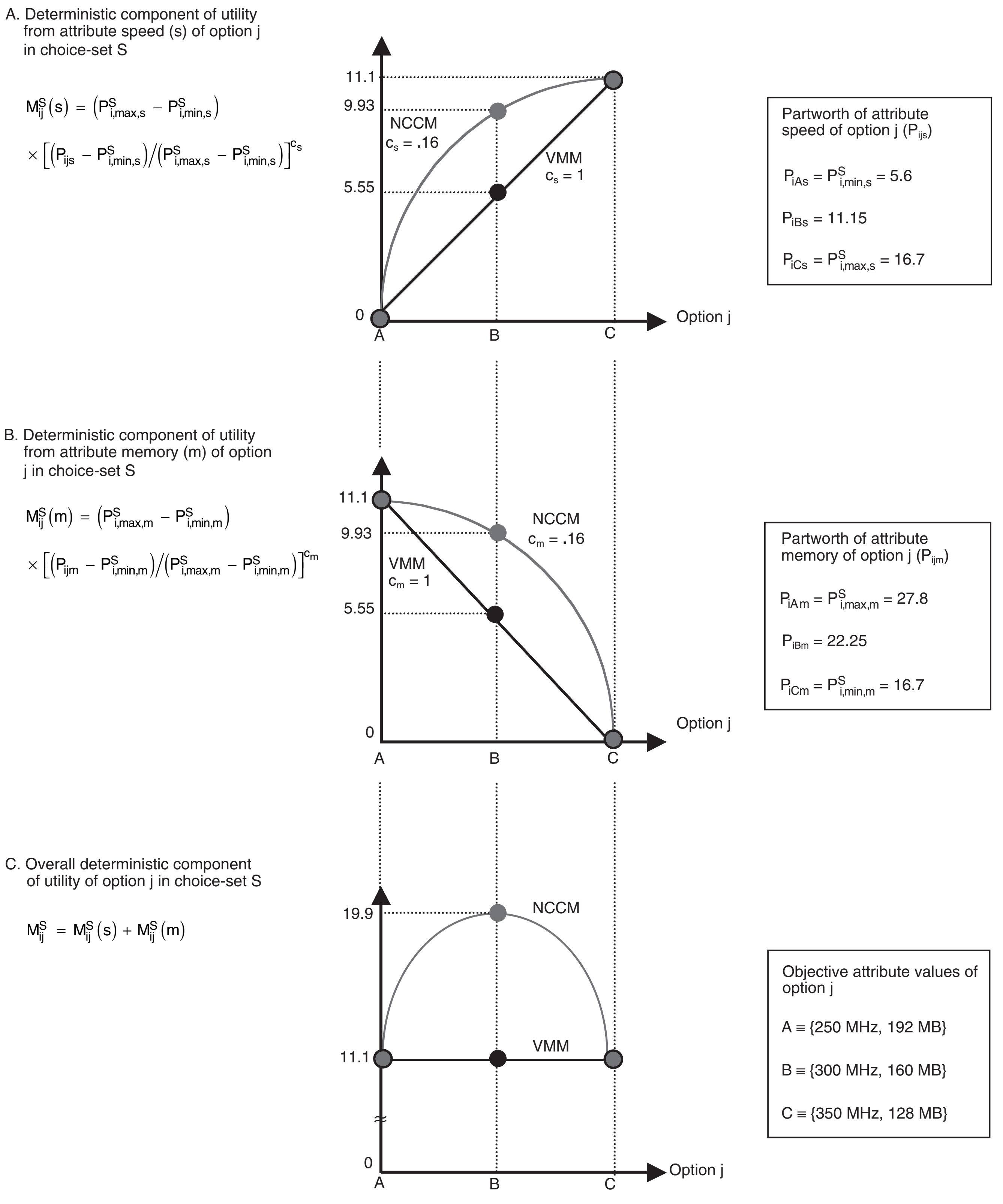
where $j, h \in S$. The parameters for estimation in this relative advantage model (RAM) are the logit parameters b and q, the loss-aversion parameters L_k, and the power parameters ψ_k (in general, there are 2d + 2 parameters, where d is the number of dimensions). Note that when q = 0, the RAM reduces to the standard value maximization multinomial logit. However, we expect that the parameters b and q will be positive, thus capturing the positive effect of higher utility and greater relative advantages on choice, respectively. In addition, by examining the square-bracketed expression in Equation 6, we expect that the loss-aversion parameters L_k will be greater than 0, thus capturing the notion that disadvantages loom larger than advantages (e.g., Simonson and Tversky 1992). The loss-aversion parameters are allowed to vary by attribute, which is consistent with evidence that the magnitude of loss aversion may differ across attributes (e.g., Dhar and Wertenbroch 2000; Hardie, Johnson, and Fader 1993; Heath et al. 2000; Tversky and Kahneman 1991; Viscusi, Magat, and Huber 1987). We also expect that the power parameters ψ_k will be greater than 1 (which satisfies the convexity assumption), and we employ attribute-specific power parameters to allow for situations in which the strength of the compromise effect varies across attributes (see Tversky and Simonson 1993).

Although Tversky and Simonson (1993) explain compromise using loss aversion, it is noteworthy that the key driver

[6]An alternative formulation for $D_{ik}(j, h)$ is an exponential rather than a power function; that is, $D_{ik}(j, h) = [L_k \times \exp(\psi_k A_{ik}(h, j))] \times (P_{ihk} > P_{ijk})$. We tested this alternative function, but the parameter estimates were not consistent with the theory.

Figure 4

THE NORMALIZED CONTEXTUAL CONCAVITY MODEL: ATTRIBUTE-SPECIFIC AND AGGREGATE UTILITY GRAPHS
(FOR A PARTICULAR CONSUMER WITH EQUAL ATTRIBUTE IMPORTANCES)



in the RAM is that disadvantages ($D_k(j, h)$) are convex functions of the corresponding advantages.[7] For example, without convexity in the disadvantage function (i.e., $\psi = 1$), the relative advantages of any set of options with equal context-independent utility (i.e., equal V_{ij}) would be identical even when loss aversion is assumed (i.e., $L > 0$). Next, we present a model that strictly employs loss aversion as the principal mechanism that underlies the compromise effect.

The Loss-Aversion Model

Two major principles that have emerged from behavioral decision research are that consumers evaluate attribute values on the basis of their deviation from a reference point (i.e., reference dependence) and that such deviations have greater impact when consumers perceive them as losses rather than gains (i.e., loss aversion) (Kahneman and Tversky 1979). These principles have also been suggested as empirical generalizations in marketing (e.g., Kalyanaram and Winer 1995; Meyer and Johnson 1995; cf. Bell and Lattin 2000). We suggest that a multiattribute choice model that incorporates reference dependence and loss aversion can capture the compromise effect.

In particular, we model choice as if each alternative were evaluated relative to a reference point defined using the midpoint of the range of objective attribute levels observed in the local choice set. That is, consistent with Tversky and Kahneman's (1991) reference-dependent model, we assume that all alternatives in the local choice set are evaluated relative to a *single* reference point R. This reference point is context dependent, because the midpoint of the attribute range is determined on the basis of the particular alternatives observed in the local choice set. The use of a reference point that is not necessarily an existing choice option is consistent with a great deal of research on reference dependence (e.g., Heath, Larrick, and Wu 1999; Kahneman and Tversky 1979; Kivetz 2003; Thaler 1985; Winer 1986).

Thus, building on Tversky and Kahneman's (1991) reference-dependent model, we suggest that the compromise effect can be captured by the following loss-aversion model (LAM):

$$(8) \qquad M_{ij}^S = \sum_k \left[\left(P_{ijk} - P_{iRk}^S\right) \times 1\left(P_{ijk} \geq P_{iRk}^S\right)\right] + \left[\lambda_k \times \left(P_{ijk} - P_{iRk}^S\right) \times 1\left(P_{ijk} < P_{iRk}^S\right)\right],$$

where

λ_k = the loss-aversion parameter of attribute k;
R = the reference option, defined as the midpoint of each attribute's observable range in the local choice set; and
P_{iRk}^S = the partworth of attribute k at the reference point (R) in choice-set S for consumer i.

According to the LAM, the probability that consumer i will choose alternative j in the context of S has the multinomial logit form (see Equation 2). The parameters for estimation are the logit scale parameter b, which we expect to be positive, and the loss-aversion parameters λ_k, which we expect to be greater than 1 (in general, there are d + 1 parameters, where d is the number of dimensions). When $\lambda_k = 1$ for all k, the LAM reduces to the standard value-maximization multinomial logit. This is true because we have $M_{ij}^S = \sum_k(P_{ijk} - P_{iRk}^S) = \sum_k P_{ijk} - \sum_k P_{iRk}^S$, which equals the commonly used partworth additive utility function minus a constant. (The additive constant given by the second term does not affect the logit choice probabilities.)

The LAM assumes that the marginal utility of an increase in an attribute level is greater below the attribute's midrange than above it. The model employs constant loss aversion, and the parameter λ_k can be interpreted as the coefficient of loss aversion for attribute k (see Tversky and Kahneman 1991, pp. 1050–51). Moreover, the LAM can capture situations in which the magnitude of the compromise effect varies across attributes by allowing the loss-aversion parameters to differ by attribute.

To illustrate how the LAM incorporates the compromise effect, we employ attribute-specific and aggregate utility graphs and use the previous example of three portable PCs (see Figure 5). The (45 degree) diagonal lines shown in Panels A and B represent the VMM and are based on Equation 8; the graphs assume that $\lambda_s = \lambda_m = 1$. Conversely, the "kinked" LAM graphs capture compromise by $\lambda_s, \lambda_m > 1$.

Panels A and B of Figure 5 show that, compared with the VMM, the LAM penalizes alternatives with lower partworths compared with the (context-dependent) reference point R. Accordingly, as is shown in the aggregate utility graph, value maximization implies that A ~ B ~ C, whereas the LAM implies that $A \prec B \succ C$. More generally, the LAM suggests that the gain in utility due to a positive deviation (from R) on one attribute typically does not suffice for the loss of utility due to the corresponding negative deviation on another attribute.

Next, we report two empirical applications that we used to compare the alternative models with one another, with the VMM, and with another (stronger) context-independent (naive) model. The first application employs the common design for investigating the compromise effect, namely, trinary choice sets defined on two attributes. The second empirical application is intended to provide a stronger and more general test of both the compromise effect and the proposed models by using larger sets of options and attributes.

EMPIRICAL APPLICATION 1: TWO-DIMENSIONAL TRIPLETS

In this empirical application, we employ a partworth function preference model (see, e.g., Green and Srinivasan 1978, 1990) and determine the individual-level partworths using the self-explicated approach (e.g., Srinivasan and Park 1997). The self-explicated task covers the entire range of attribute levels used in the choice study; thus, the self-explicated partworths are independent of the local choice context. We briefly describe this approach and subsequently detail the method we used to obtain the experimental choice data and the self-explicated partworths. We then report the calibration and validation of the alternative models using these data.

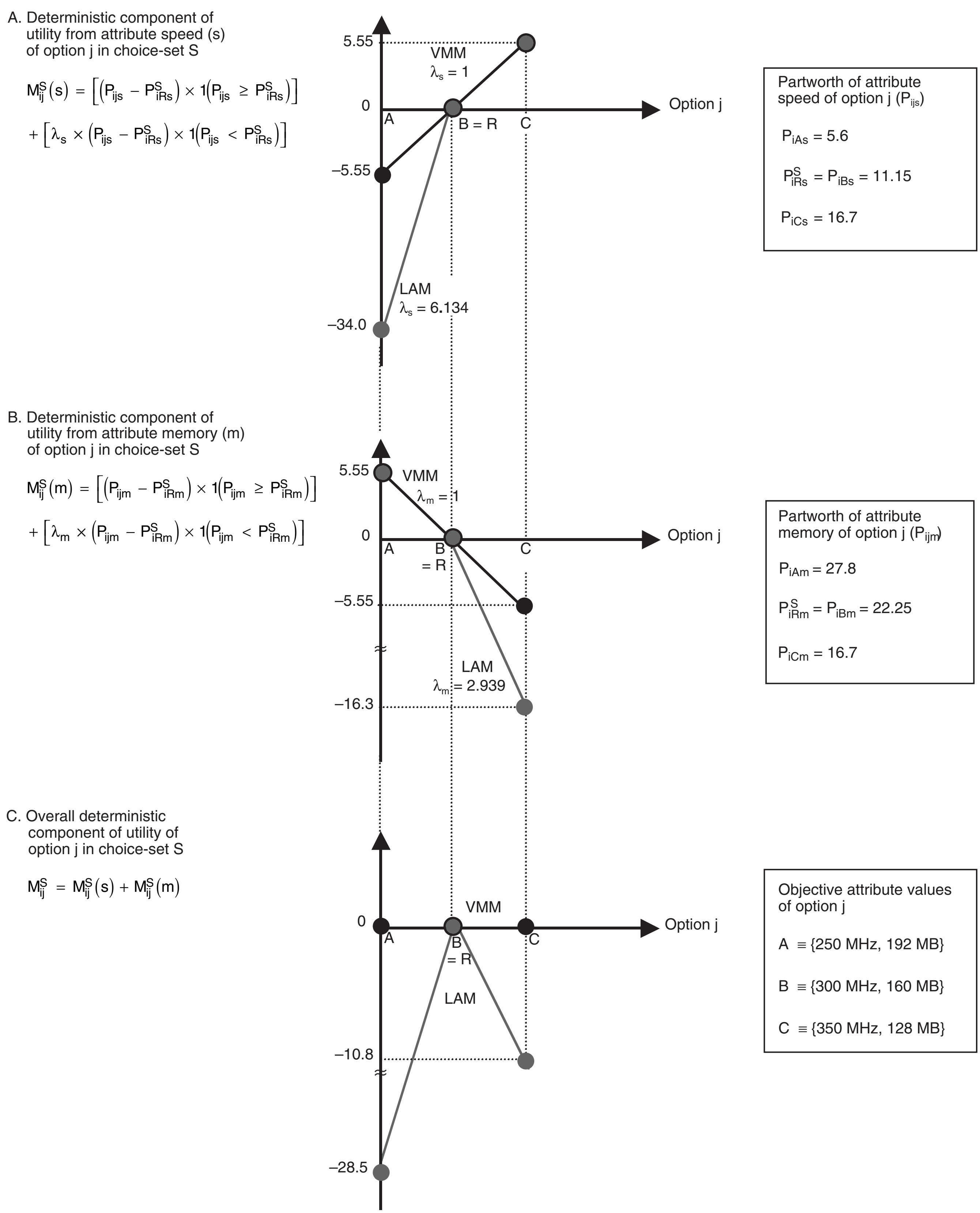
Self-Explicated Partworths

There is growing evidence on the robust validity of measuring attribute partworths with the self-explicated approach (Leigh, MacKay, and Summers 1984; Srinivasan 1988; Srinivasan and Park 1997). We used a self-explicated proce-

[7]A detailed proof for this proposition is available from the authors.

Figure 5

THE LOSS-AVERSION MODEL: ATTRIBUTE-SPECIFIC AND AGGREGATE UTILITY GRAPHS
(FOR A PARTICULAR CONSUMER WITH EQUAL ATTRIBUTE IMPORTANCES



dure based on the work of Srinivasan (1988).[8] According to this approach, there are two stages in the data collection of the self-explicated partworths: (1) rating the desirability of each attribute level used in the study and (2) indicating the relative importance of each attribute. The individual-level self-explicated partworths for the various attribute levels are then calculated by multiplying the attribute importances by the desirability ratings. For each respondent, the ranges (i.e., maximum – minimum) of the partworth functions sum to 100 over the two attributes, and the range is proportional to the importance rating. Individual-level analysis indicates that participants vary greatly with respect to the shape of their partworth functions (e.g., concave, convex, linear). The finding that the individual-level partworth functions vary widely across participants highlights the advantage of using an individual-level partworth function preference model rather than less flexible models, such as the (linear) vector model or the ideal-point model (see Green and Srinivasan 1978, 1990). Next, we describe a questionnaire-based study in which we obtained the (context-independent) self-explicated partworths and the choice data used to calibrate and validate the alternative models.

Method

Participants. The participants were 1088 travelers who were waiting for their flights at domestic terminals in a major airport. They were between 18 and 70 years of age and represented a wide range of demographic characteristics.

Choice stimuli. We used two product categories that most airport travelers are familiar with: portable PCs and speakers. For each attribute, the introduction to all tasks (self-explicated and choice) specified the "range of typical attribute values offered in the marketplace," which is consistent with the finding of Assar and Chakravarti (1984) that attribute range knowledge enables respondents to comprehend better brand-attribute information and to make meaningful attribute trade-offs. Each product category included two attributes with five levels each (in addition to the two boundary levels that marked the typical market range of an attribute); the attribute levels were also used in the self-explicated task. Accordingly, in each category, there were five Pareto-optimal choice options that were derived from a linear attribute trade-off function, similar to the design of choice stimuli in previous studies of the compromise effect. The attribute levels and ranges were based on the typical values found in the market at the time of the data collection. Given the evidence that responses to price may differ qualitatively from responses to other attributes (e.g., Hardie, Johnson, and Fader 1993; Simonson and Tversky 1992), we intentionally included price as an attribute in the speakers category but not in the portable PC category. Table 1 describes the five product options used in the portable PC and speakers categories. In each category, the five alternatives served as the basis for three different trinary choice sets, as was previously shown in Figure 1.[9]

Procedure and design. Participants were randomly assigned to one of six major conditions in a 2 (choice type: calibration versus validation) × 3 (choice set: 1 versus 2 versus 3) between-subjects design. Each respondent provided information on both product categories. Following their choice (in each product category) in one of the three choice sets, the calibration participants completed the two stages of the self-explicated task described previously (we used this order of tasks to ensure that calibration choices were not biased by the self-explicated task). Regardless of their choice-set condition, all the calibration participants faced the same self-explicated task that covered the entire range of attribute levels used in the choice study. The validation sample participants also made one trinary choice in each category, but they did not complete the self-explicated task. Participants in all conditions were instructed not to

[8]A detailed description of the procedure used in this application is available from the authors.

[9]For the speaker category, the attribute v of Figure 1 represents the negative of price so that larger v values denote less expensive options.

Table 1
CHOICE OPTIONS AND SHARES (EMPIRICAL APPLICATION 1)

			Share (%)					
			Calibration Sample			*Validation Sample*		
Option/Attribute	*Speed in MHz (w)*	*Memory in MB (v)*	*Set 1 (n = 151)*	*Set 2 (n = 148)*	*Set 3 (n = 164)*	*Set 1 (n = 205)*	*Set 2 (n = 200)*	*Set 3 (n = 220)*
Portable PC								
A	250	192	6	—	—	13	—	—
B	300	160	50	18	—	49	15	—
C	350	128	44	51	24	38	51	25
D	400	96	—	31	47	—	35	47
E	450	64	—	—	29	—	—	28
			Calibration Sample			*Validation Sample*		
	Power in Watts (w)	*Price in US$ (v)*	*Set 1 (n = 119)*	*Set 2 (n = 107)*	*Set 3 (n = 125)*	*Set 1 (n = 206)*	*Set 2 (n = 199)*	*Set 3 (n = 216)*
Speakers								
A	50	100	9	—	—	17	—	—
B	75	130	45	15	—	40	21	—
C	100	160	46	46	26	44	44	29
D	125	190	—	39	46	—	35	45
E	150	220	—	—	28	—	—	26

look back at their previous responses (for a description of the design and the sample sizes in each condition, see Table 2).

It is important to note that in all conditions, we counterbalanced both the order of product categories and the positions of choice options on the page (see Drolet 2002; Huber et al. 1993). However, because we did not find any significant order or position effects, we subsequently aggregated the results across the order and position subconditions.

Results

Table 1 reports the choice shares for the portable PC and speakers categories in the calibration and validation samples. Options were relatively more attractive when they were in the middle than when they were extreme. As Simonson and Tversky (1992, p. 290) suggest, the compromise effect can be measured by statistically testing whether $P_A(B; C) > P_D(B; C)$ and whether $P_B(C; D) > P_E(C; D)$. For example, in the calibration sample of the portable PC category, the share of B relative to C is 53% in set 1 (i.e., .5/[.5 + .44]) versus 26% in set 2 (i.e., .18/[.18 + .51]), a difference of 27% ($z = 4.2$; $p < .001$). The "Calibration" and "Validation" columns in Table 3 show statistically significant compromise effects in all eight possible tests (i.e., two compromise effect measures in two categories and two samples); the measures of compromise ranged from 15% to 34% (mean = 24%; median = 25%).

Calibration and Validation of Models

We subsequently report the results of the calibration (estimation) and validation of the alternative choice models that are designed to capture the compromise effect. We begin by estimating the models and considering the interpretation of their parameters. Subsequently, we compare the models with one another, with the VMM, and with a stronger context-independent (naive) model that adjusts for

Table 2

EXPERIMENTAL DESIGN AND SAMPLE SIZES

Condition	*Part 1 (Validation Sample: Choice Task)*	*Part 1 (Calibration Sample: Choice Task)*	*Part 2 (Calibration Sample: Self-Explicated Tasks)*
Condition 1 $N_{pc} = 220$ $N_{speakers} = 216$	Set 3	—	—
Condition 2 $N_{pc} = 164$ $N_{speakers} = 125$	—	Set 3	Full range of attribute levels
Condition 3 $N_{pc} = 200$ $N_{speakers} = 199$	Set 2	—	—
Condition 4 $N_{pc} = 148$ $N_{speakers} = 107$	—	Set 2	Full range of attribute levels
Condition 5 $N_{pc} = 205$ $N_{speakers} = 206$	Set 1	—	—
Condition 6 $N_{pc} = 151$ $N_{speakers} = 119$	—	Set 1	Full range of attribute levels

Notes: For definitions of Sets 1, 2, and 3, see Figure 1. Fewer participants had detailed familiarity with the speakers product category in order to provide the information required for the self-explicated task; thus, in general, the sample sizes in the calibration conditions are lower for the speakers category than for the portable PC category.

Table 3

PREDICTED AND OBSERVED COMPROMISE EFFECT MEASURES (EMPIRICAL APPLICATION 1)

Category	*VMM*	*GCM*	*CCM*	*NCCM*	*LAM*	*RAM*	*Calibration*	*Validation*
Portable PCs								
$P_A(B, C) – P_D(B, C)$	–1%	1%	20%	20%	22%	3%	27%	34%
$P_B(C, D) – P_E(C, D)$	–3%	0%	26%	21%	30%	1%	28%	25%
Average compromise measure MAD	32%	29%	8%	9%	9%	28%	—	—
Speakers								
$P_A(B, C) – P_D(B, C)$	–7%	–5%	20%	19%	19%	–4%	25%	15%
$P_B(C, D) – P_E(C, D)$	–5%	0%	23%	18%	21%	–2%	18%	17%
Average compromise measure MAD	22%	19%	6%	3%	4%	19%	—	—
Overall Model MAD	26.8%	24.0%	6.5%	5.8%	6.3%	23.3%	—	—

Notes: All eight observed compromise measures in the calibration and the validation samples are significantly greater than 0 (all *p*s < .01).

possible biases in the measurement of partworths; we compare the models in terms of their predictive validity, fit, and ability to capture the observed compromise effects.

Estimation of models. Using maximum likelihood estimation, we estimated the parameters of the four alternative models (and the two additional benchmark models) using the full calibration sample.[10] We calibrated the models separately for the portable PC and speakers product categories.[11] For all the models, in general, the estimation results were consistent across the two categories.

CCM (Equations 1 and 2). As is shown in Table 4, the logit scale parameter b and the concavity parameters c_w and c_v are positive and significant. Furthermore, consistent with the notion of a contextual concavity in the utility function, both concavity parameters are significantly smaller than 1 ($p < .001$). We also tested for the possibility of restricting $c_w = c_v$. However, based on the likelihood ratio test, the improvement from relaxing this restriction was statistically significant at the 1% level.[12] In particular, in the portable PC category, the attribute memory was significantly more concave than the attribute speed (i.e., $c_{memory} < c_{speed}$; $p < .01$). In the speakers category, consistent with the notion that consumers avoid the lowest-price, lowest-quality option (Simonson and Tversky 1992), the attribute price was significantly more concave than the attribute power (i.e., $c_{price} < c_{power}$; $p < .01$). Therefore, the subsequent tests of fit and predictive validity for the CCM employ attribute-specific concavity parameters.

NCCM (Equations 2 and 3). We found that using a restricted NCCM with $c_w = c_v = c$ did not result in a significant loss of fit ($p > .1$). We therefore proceeded with the more parsimonious NCCM. Table 4 shows that the logit scale parameter b and the concavity parameter c are positive and significant. Furthermore, consistent with contextual concavity, the c parameter is significantly smaller than 1 ($p < .001$).

RAM (Equations 4–7). Because of the high nonlinearity in the estimated disadvantage functions (Equation 6), we needed to restrict $\psi_w = \psi_v = \psi$ and employ a grid search over the values of ψ to find an optimal estimate of $\psi = 5$. We also tested the restriction $L_w = L_v$, but it resulted in a significant loss of fit ($p < .01$). Specifically, in the portable PC category, the attribute speed exhibited greater loss aversion than did the attribute memory (i.e., $L_{speed} > L_{memory}$; $p < .01$). In the speakers category, consistent with prior research (see Hardie, Johnson, and Fader 1993; Simonson and Tversky 1992), the attribute power exhibited more loss

[10]We performed the maximum likelihood estimation using the TSP statistical package (Hall and Cummins 1999).

[11]Nine participants from the calibration sample in the portable PC category and six in the speakers category were ignored because they had nonincreasing partworth functions with respect to the attribute levels.

[12]For nested models, significance is determined according to the likelihood ratio test: (2LL[unrestricted model]) – (2LL[restricted model]) ~ $\chi^2(\Delta r)$, where Δr is the difference in the number of parameters between the unrestricted and restricted models.

Table 4

MODEL ESTIMATES USING THE FULL CALIBRATION SAMPLE IN EMPIRICAL APPLICATION 1

	Portable PCs (N = 454)				*Speakers (N = 345)*		
	Estimate	*Standard Error*	*p-Value*		*Estimate*	*Standard Error*	*p-Value*
CCM Estimates[a]							
b	.387	.084	.000	b	.446	.088	.000
c_{speed}	.474	.067	.000	c_{power}	.574	.055	.000
c_{memory}	.353	.086	.000	c_{price}	.378	.068	.000
–2LL	920.4			–2LL	638.7		
NCCM Estimates							
b	.051	.007	.000	b	.064	.009	.000
c	.160	.095	.000	c	.315	.098	.000
–2LL	914.8			–2LL	673.5		
RAM Estimates[b]							
b	.022	.009	.011	b	.031	.009	.000
q	.673	.196	.001	q	1.304	.262	.000
L_{speed}	.015	.033	.650	L_{power}	1.745	4.664	.708
L_{memory}	.181e-03	.274e-03	.510	L_{price}	.268e-03	.267e-03	.316
–2LL	941.1			–2LL	667.0		
LAM Estimates[c]							
b	.020	.007	.008	b	.040	.008	.000
λ_{speed}	6.134	2.170	.009	λ_{power}	4.161	.853	.000
λ_{memory}	2.939	1.048	.032	λ_{price}	1.605	.397	.064
–2LL	905.6			–2LL	654.5		
GCM Estimates							
b	.100	.069	.150	b	.469	.188	.013
c_{speed}	.824	.221	.213	c_{power}	.621	.086	.000
c_{memory}	.716	.216	.095	c_{price}	.471	.088	.000
–2LL	949.6			–2LL	657.5		

[a]For the parameters c_w and c_v, the meaningful null hypothesis is the VMM ($c_w = 1$ and $c_v = 1$). Thus, the *p*-values reported for these parameters in the CCM, NCCM, and GCM are with respect to $c_w \geq 1$ and $c_v \geq 1$.

[b]$\psi_{speed} = \psi_{memory} = \psi = 5$ for PCs, and $\psi_{power} = \psi_{price} = \psi = 5$ for speakers.

[c]For the parameters λ_w and λ_v, the meaningful null hypothesis is the VMM ($\lambda_w = 1$ and $\lambda_v = 1$). Thus, the *p*-values reported for these parameters in the LAM are with respect to $\lambda_w \leq 1$ and $\lambda_v \leq 1$.

aversion than did the attribute price (i.e., $L_{power} > L_{price}$; $p < .01$). Furthermore, Table 4 indicates that all the estimated RAM parameters (b, q, L_w, L_v) are positive, as we expected. It is noteworthy that the estimated RAM satisfies the modeling requirements that Tversky and Simonson (1993) set for capturing the compromise effect, albeit the L parameters are not statistically significant. In particular, disadvantages both outweigh and grow faster than corresponding advantages.

LAM (Equations 2 and 8). Table 4 shows that for LAM estimates, the logit scale parameter b and the loss-aversion parameters λ_w and λ_v are positive and significant. Furthermore, as we predicted, all loss-aversion parameters are greater than 1 (all *p*s < .05, except for λ_{price}, for which $p < .07$). Averaging across attributes, the loss-aversion coefficient λ was approximately 4.5 in the PC category and 2.9 in the speakers category. However, given that the restriction $\lambda_w = \lambda_v$ resulted in significant loss in fit ($p < .01$), we proceeded with attribute-specific loss-aversion parameters. More specifically, the attribute speed exhibited greater loss aversion than did the attribute memory (i.e., $\lambda_{speed} > \lambda_{memory}$; $p < .01$), and the attribute power exhibited greater loss aversion than did the attribute price (i.e., $\lambda_{power} > \lambda_{price}$; $p < .01$).[13]

VMM. To test whether incorporating the compromise effect improves fit and predictive validity compared with the standard VMM, we also estimated the logit scale parameter b of the VMM in Equation 9. This scale parameter is positive and significant, both in the PC category (b = .04; $p < .001$; –2 log-likelihood [LL] = 956.9) and in the speakers category (b = .05; $p < .001$; –2LL = 698.6).

$$\text{(9)} \qquad \Pr_i(j) = \frac{\exp(bM_{ij})}{\sum_h \exp(bM_{ih})},$$

where $M_{ij} = \Sigma_k(P_{ijk})$, and $j, h \in S$.

Global concavity model. We also compared the alternative models with a stronger context-independent (naive) model than the VMM. The global concavity model (GCM) is similar to the CCM in that it induces additional concavity (i.e., diminishing sensitivity) on the estimated (context-independent) partworth functions (see Equation 10). That is, regardless of the original shape of a measured partworth function (which transforms objective attribute values into subjective utilities), the GCM adds another layer of concavity based on the choice data. However, unlike the CCM, this additional concavity is applied globally (i.e., independent of the choice set) to all attribute levels used in the study, such that the modeled utility of an option does not vary across choice sets. The GCM, then, is a naive model because it does not account for the impact of the local choice set, and it is not expected to capture the compromise effect.

$$\text{(10)} \qquad M_{ij} = \sum_k \left(P_{ijk}\right)^{c_k}, \text{ and } j \in S.$$

The GCM is a stronger model than the VMM because it allows for adjustments in the estimated partworths, which may be needed because of possible measurement biases. Use of the GCM also allows for partitioning of the context-free and context-dependent effects of the various models. In particular, any superior predictive ability of the alternative models over the VMM might be attributed to adjustments of stated preferences (obtained in the partworth estimation phase) to observed choices (i.e., revealed preferences). However, any such effects should also aid the GCM, which can correct the measured utilities. Thus, better performance of the proposed models over the (context-independent) GCM would indicate the importance of the models being context specific.

We estimated the parameters of the GCM on the basis of Equation 10 and a multinomial logit form similar to that of the VMM (see Equation 9). As we expected, the GCM's logit scale parameter b and concavity parameters c_w and c_v are positive (see Table 4). Furthermore, consistent with the notion that the GCM can adjust the measured (stated) partworths by inducing global concavity, the concavity parameters are significantly smaller than 1 in the speakers category (both *p*s < .001) but not in the PC category. We also tested the restriction $c_w = c_v$, but the improvement from relaxing this restriction is statistically significant at the 1% level, and therefore we calibrate and validate the GCM using attribute-specific concavity parameters.

Predictive validity and fit. We used several measures to compare the predictive validity and fit of the competing models, which we discuss next.

Aggregate-level predictive validity. We measured the aggregate-level predictive validity of the alternative models using choice data from two choice-set calibration conditions and one choice-set validation condition. For example, we used the self-explicated partworths and observed choices of the (calibration) participants assigned to sets 2 and 3 to calibrate the models; we then used the estimated parameters to predict the choices of the (validation) participants assigned to set 1. That is, we compared the average predicted logit choice probabilities for set 1 (based on calibration sets 2 and 3) with the choice proportions obtained in the validation sample for set 1.[14] Thus, the procedure consists of a cross–choice set, cross-samples validation. We applied the same aggregate-level validation procedure to predict choice sets 2 and 3.

As a measure of the aggregate-level predictive validity, we used the mean absolute deviation (MAD) between the predicted choice shares (based on the calibration sample) and the observed validation choice shares. Table 5 shows that, averaged across the three choice sets and in both product categories, the CCM, NCCM, and LAM have superior MAD measures than the VMM, GCM, and RAM. Furthermore, the two contextual concavity models have improved predictive validity compared with the LAM, and the NCCM is somewhat better than the CCM.

Because this is a cross-samples validation, we expected some sampling error to result from the calibration being based on a different sample of respondents than the validation sample (see Huber et al. 1993). To estimate the degree of sampling error, we used a bootstrap procedure with 100 replications; each replication consisted of n pseudorespondents (n is the number of respondents summed over the two

[13]This result is consistent with the previous result for the CCM that customers avoided the lowest-price, lowest-quality option.

[14]In terms of the experimental design in Table 2, in this case we calibrated the model on the basis of experimental conditions 2 and 4 to predict the results in condition 5.

Table 5
COMPARISON OF MODELS' FIT AND PREDICTIVE VALIDITY
(ACROSS CHOICE SETS IN EMPIRICAL APPLICATION 1)

	VMM	*GCM*	*CCM*	*NCCM*	*LAM*	*RAM*
Portable PCs						
Aggregate-level prediction (MAD)	11.1%	10.4%	6.3%	6.4%	7.0%	10.4%
Improvement in aggregate-level prediction (MAD) over VMM (Equation 11)	—	8%	54%	52%	46%	8%
Improvement in aggregate-level prediction (MAD) over GCM	—	—	50%	48%	41%	0%
Individual-level prediction (–2LL)	960.4	965.5	947.1	920.2	956.2	956.9
Model fit (BIC)	963.0	968.0	938.8	927.1	924.0	971.7
Speakers						
Aggregate-level prediction (MAD)	9.5%	8.4%	5.9%	5.4%	6.2%	10.0%
Improvement in aggregate-level prediction (MAD) over VMM (Equation 11)	—	16%	51%	58%	47%	–7%
Improvement in aggregate-level prediction (MAD) over GCM	—	—	42%	50%	37%	–27%
Individual-level prediction (-2LL)	702.3	670.7	662.5	683.3	674.3	692.2
Model fit (BIC)	704.4	675.0	656.2	685.2	672.0	697.2
Number of parameters	1	3	3	2[a]	3	5[b]

[a]The NCCM had only two parameters because the concavity parameters did not vary by attribute (i.e., $c_w = c_v = c$).
[b]The RAM had five parameters because we restricted $\psi_w = \psi_v = \psi$.

calibration samples used to predict the shares in a validation sample). To construct a population that resembles the validation sample, each pseudorespondent was randomly drawn from the validation sample with replacement. We obtained the MAD for a replication by calculating the choice proportions in the pseudosample and by comparing them with the actual validation choice proportions. We averaged the bootstrap MADs over the 100 replications. We repeated this entire process for the three validation sets for each product category. The bootstrap average MAD estimates (averaged across choice sets) for the PC and the speakers categories were 2.1% and 2.4%, respectively. On the basis of these estimates, we calculated the improvement in predictive validity obtained by using each of the alternative models rather than the standard VMM (after accounting for the sampling error-based MAD) as follows:

$$(11)\qquad \text{Improvement in predictive validity} = \frac{[\text{MAD}(\text{VMM}) - \text{MAD}(\text{alternative model})]}{[\text{MAD}(\text{VMM}) - \text{MAD}(\text{bootstrap sampling})]}.$$

As Table 5 shows, the contextual concavity models and the LAM provided a substantial improvement (46%–58%) over the VMM, whereas the GCM and RAM did not. Using Equation 11 (but with VMM replaced by GCM), we also found that the contextual concavity models and the LAM provided large improvements (37%–50%) in predictive validity over the context-independent GCM. The results convincingly show that accounting for the (local) choice context can significantly improve aggregate predictive validity. In addition, the two contextual concavity models yielded greater improvements than the LAM.

Individual-level predictive validity. We estimated the models' parameters using two calibration choice-set conditions, and using the self-explicated partworths obtained in a third calibration choice-set condition, we predicted the choice probabilities in this third set. In contrast to the aggregate-level predictions, because the calibration sample includes partworth estimates and choices at the individual level, we can calculate the –2LL individual-level measure of predictive validity (smaller values indicate better prediction ability). However, a confounding aspect of these predictions, which the aggregate-level predictions do not suffer, is that the choice made in the calibration set may have affected the partworths elicited in the subsequent self-explicated task. Table 5 shows that, pooled across the three choice sets and two product categories, the individual-level measure of predictive validity favored the contextual concavity models and the LAM over the VMM and RAM. In addition, the three leading models (i.e., CCM, NCCM, and LAM) outperformed the GCM in the PC category but not in the speakers category. Overall, this pattern is consistent with the results of the aggregate-level predictions.

Model fit. We also compared the models' fit using the entire calibration sample. To penalize for the number of parameters, we employed the Schwarz Bayesian information criterion (BIC) measure (smaller values indicate better fit). As is shown in Table 5, the pattern of the BIC results is consistent with the validation findings.

Recall that the VMM is nested in all five models. Thus, we employed the likelihood ratio test described previously to compare the alternative models with the VMM. The likelihood ratio test indicated that all the models provided a significant improvement in fit over the VMM, though the improvements were most pronounced for the three leading context-dependent models.

Additional evidence. To illustrate the superior predictive validity of the CCM, NCCM, and LAM and to demonstrate their ability to predict the compromise effect, in Table 6 we report the aggregate-level predictions of the six models and the actual observed choices in choice-set 3 of the PC category.[15] The three leading models were better able to predict

[15]To conserve space, we report the predicted versus observed choices for only one of the six choice sets (2 product categories × 3 choice sets). Overall, the CCM, NCCM, and LAM dominate the VMM, GCM, and RAM

Table 6

AGGREGATE-LEVEL PREDICTIONS AND VALIDATION CHOICE SHARES FOR CHOICE SET 3 IN THE PORTABLE PC CATEGORY (EMPIRICAL APPLICATION 1)

Option/Model	*VMM*	*GCM*	*CCM*	*NCCM*	*LAM*	*RAM*	*Validation Sample*
C	35%	30%	19%	28%	16%	28%	25%
D	32%	34%	45%	45%	46%	33%	47%
E	33%	36%	35%	27%	38%	39%	28%
MAD	10%	9%	5%	2%	7%	9%	—

the compromise effects observed in the choice data. Recall that the compromise effect states that $P_A(B; C) > P_D(B; C)$ and $P_B(C; D) > P_E(C; D)$. In contrast, value maximization combined with the highly plausible ranking condition actually predicts a result that is diametrically opposed to the compromise effect; that is, $P_A(B; C) < P_D(B; C)$ and $P_B(C; D) < P_E(C; D)$. Indeed, as is shown in Table 3, the VMM consistently predicted a "reverse" (or negative) compromise effect. However, the choice data obtained in the calibration and validation samples exhibited substantial compromise effects. Table 3 also reports the compromise measures derived from the aggregate-level predictions of the alternative models. Consistent with the predictive validity and fit measures reported previously, the CCM, NCCM, and LAM captured the observed compromise effects quite well, whereas the other models did not. More specifically, we calculated the MAD between the predicted compromise measures (based on the calibration sample) and the compromise measures observed in the validation sample. We averaged the MAD across the two compromise measures—that is, $P_A(B, C) – P_D(B, C)$ and $P_B(C, D) – P_E(C, D)$—and we report it separately for each alternative model and product category. The MAD compromise measures were consistently much better for the CCM, NCCM, and LAM than for the VMM, GCM, and RAM. It is noteworthy that the overall superiority of the three leading models to the GCM supports the notion that these models capture context effects rather than only adjust for discrepancies between stated (partworth) utilities and revealed preferences (based on choices).

EMPIRICAL APPLICATION 2: LARGER SETS OF OPTIONS AND ATTRIBUTES

The previous empirical application tested the alternative models in a context similar to previous demonstrations of the compromise effect. A question that naturally arises is whether the effect survives in choice sets with more than three options and two dimensions. Relatedly, it is important to test the proposed models compared with the context-independent VMM and GCM in a more complex setting that is closer, in terms of task dimensionality, to a typical conjoint analysis study. The present empirical application addresses these issues by employing choice sets with five alternatives defined on four attributes. In addition, instead of measuring partworths using the self-explicated approach, we employ the commonly used full-profile conjoint analysis technique (see, e.g., Green and Srinivasan 1990). The conjoint analysis task covers the entire range of attribute levels used in the choice study and therefore estimates context-independent partworths. We briefly describe the conjoint analysis used and then report the method and results that pertain to the choice data and the calibration and validation of the models.

Conjoint Analysis Partworths

To construct the stimulus set of the full-profile conjoint analysis, we used a fractional factorial design with three levels of each of four attributes (i.e., the conjoint analysis task used the two extreme levels and an intermediate level of each attribute used in the choice study). We created a set of 18 cards, and each participant in the conjoint task was asked to rank-order the 18 profiles. We estimated each respondent's partworths using Conjoint Linmap software (Bretton-Clark 1989). An analysis of the individual-level partworth functions revealed that the shape of the partworth functions (e.g., concave, convex, linear) varied greatly across participants. Next, we describe a lab study in which we obtained the (context-independent) partworths and choice data required for calibrating and validating the alternative models.

Method

Participants. The participants were 205 students at a private West Coast university.[16] They were paid $7 each for their participation in the study, which took place in a behavioral research lab.

Choice stimuli. We used a product category that university students are familiar with, namely, portable PCs. The introduction to both the conjoint analysis and the choice tasks specified, for each attribute, the "range of typical attribute values offered in the marketplace." There were four attributes with six levels each (in addition to the two boundary levels that marked the typical market range of an attribute). Accordingly, there were six Pareto-optimal choice alternatives, which we used to construct two different choice sets, each of which included five portable PCs (see Table 7). We based the attribute levels and ranges on the typical values found in the market at the time of data collection. Respondents were told that all the options were identical on all other attributes, including price.

Procedure and design. Participants were randomly assigned to one of two conditions: either choice-set 1 (n = 101) or choice-set 2 (n = 97). We counterbalanced, between subjects, the positions of choice options on the page (there were no significant position effects). After making a choice in one of the two sets, participants received an unrelated filler task (that took about ten minutes), before completing the conjoint analysis card-sorting. Regardless of their choice-set condition, all participants faced the same con-

across the six choice sets. The predicted shares for the remaining five sets are available from the authors.

[16]Of 205 participants, 7 (i.e., 3%) had inconsistent preferences (greater than 15% violations while fitting the data) in the conjoint task; thus, we dropped them from subsequent analyses.

Table 7
CHOICE OPTIONS AND SHARES IN THE PORTABLE PC CATEGORY (EMPIRICAL APPLICATION 2)

					Share (%)	
Option/Attribute	*Memory (MB)*	*Hard Drive (GB)*	*Speed (MHz)*	*Battery Life (Hours)*	*Set 1 (n = 101)*	*Set 2 (n = 97)*
A	128	15	2000	4.5	3	—
B	256	20	1900	4.0	24	9
C	384	25	1800	3.5	32	26
D	512	30	1700	3.0	32	42
E	640	35	1600	2.5	9	13
F	768	40	1500	2.0	—	9

joint analysis task, which covered the entire range of attribute levels used in the choice study.

Results

Table 7 reports the choice shares in both sets of portable PCs. Options were more attractive when they were closer to the middle than to the extreme ends of the choice set. According to the notion of extremeness aversion (see Simonson and Tversky 1992), the compromise effect can be statistically tested in this design by means of five different contrasts, each between the relative shares of two options that exist in both choice sets. The five measures capture the extent to which options lose (gain) relative share when they become (move away from being) the extreme option in the choice set. For example, the share of B relative to D is 43% (.24/[.24 + .32]) when both options are intermediate (in set 1), but the share of B relative to D drops to 18% (.09/[.09 + .42]) when B becomes an extreme option (in set 2), thus creating a difference of 25% ($z = 2.764$; $p < .01$). The rightmost column of Table 8 shows that the results were in the predicted direction in all five possible tests of the compromise effect and were statistically significant in four of them; the measures of compromise ranged from 2% to 32% (mean = 17%; median = 16%). Next, we report the results of the calibration and validation of the alternative choice models.

Calibration and Validation of Models

Estimation of models. Using the data and maximum likelihood estimation, we calibrated the four alternative models and the benchmark (context-independent) VMM and GCM. We estimated a parsimonious version of each model, which involved model parameters that did not vary across the four attributes. Given that in this section we focus on applying the models to expanded dimensional spaces, it is important to validate parsimonious models because the use of attribute-specific parameters becomes prohibitively expensive (i.e., in terms of degrees of freedom) as the number of attributes increases. As is shown in Table 9, the estimated parameters of all models are consistent with the underlying theoretical motivation. For example, the contextual concavity parameters of the CCM and NCCM are significantly smaller than 1 (both ps < .001), whereas the loss-aversion parameter of the LAM is greater than 1 (though it did not approach statistical significance). It is important to note that we also calibrated and tested an attribute-specific GCM, but the predictive validity and fit comparisons remained the same even when we tested this more flexible model against our restricted (context-dependent) models.

Predictive validity and fit. To compare the alternative models, we employed the measures of predictive validity and fit reported previously.

Aggregate-level predictive validity. The aggregate-level predictions consisted of a cross–choice set, cross-samples validation. Specifically, we calibrated the models by using the conjoint partworths and the observed choices of participants assigned to one set; we then predicted the choices of participants assigned to the other set. For example, we compared the average predicted logit choice probabilities for set 1 (based on a calibration using set 2) with the choice proportions actually observed in set 1. We applied the same aggregate-level validation procedure to the other choice set. Table 10 shows that, averaged across the two choice sets, the CCM, NCCM, and LAM have superior MAD measures than the VMM, GCM, and RAM.

We estimated the degree of sampling error using the bootstrap procedure discussed previously; the bootstrap

Table 8
PREDICTED AND OBSERVED COMPROMISE MEASURES (EMPIRICAL APPLICATION 2)

Portable PC	*VMM*	*GCM*	*CCM*	*NCCM*	*LAM*	*RAM*	*Observed Compromise Measure*[a]
$P_{A,D,E}(B, C) – P_{D,E,F}(B, C)$	–4%	–8%	16%	20%	–7%	5%	16%
$P_{A,C,E}(B, D) – P_{C,E,F}(B, D)$	–5%	–7%	18%	22%	6%	9%	25%
$P_{A,C,D}(B, E) – P_{C,D,F}(B, E)$	–7%	–6%	28%	32%	9%	9%	32%
$P_{A,B,D}(C, E) – P_{B,D,F}(C, E)$	–4%	3%	11%	12%	18%	4%	12%
$P_{A,B,C}(D, E) – P_{B,C,F}(D, E)$	–3%	2%	8%	9%	1%	–1%	2%
Average compromise measure MAD	22%	21%	4%	3%	14%	12%	—

[a]The first four (of five) observed compromise measures are significantly greater than 0 (all ps < .01).

Notes: The five measures are based on the notion of extremeness aversion (see Simonson and Tversky 1992) and capture the extent to which options lose (gain) relative share when they become (move away from being) the extreme option in the choice set. An alternative compromise measure assumes that options gain (lose) share as they become (move away from being) the middle option. This measure, which implies that $P_{A,B,E}(C, D) – P_{B,E,F}(C, D) > 0$, is equal to 12% ($p < .01$) in the observed choices.

Table 9

MODEL ESTIMATES USING THE FULL CALIBRATION SAMPLE IN EMPIRICAL APPLICATION 2 (N = 198)

	Estimate	*Standard Error*	*p-Value*
CCM Estimates[a]			
b	.301	.064	.000
c	.595	.051	.000
–2LL	533.5		
NCCM Estimates			
b	.075	.011	.000
c	.364	.071	.000
–2LL	531.2		
RAM Estimates[b]			
b	.010	.013	.476
q	.453	.118	.000
L	.54e-9	.60e-9	.370
–2LL	550.0		
LAM Estimates[c]			
b	.009	.010	.385
λ	8.817	9.823	.213
–2LL	527.8		
GCM Estimates			
b	.263	.075	.000
c	.645	.059	.000
–2LL	547.3		

[a]For the parameter c, the meaningful null hypothesis is the utility maximization model (c = 1). Thus, the *p*-values reported for this parameter in the CCM, NCCM, and GCM are with respect to the null hypothesis: $c \geq 1$.
[b]$\psi = 10$.
[c]For the parameter λ, the meaningful null hypothesis is the utility maximization model ($\lambda = 1$). Thus, the *p*-value reported for this parameter in the LAM is with respect to the null hypothesis: $\lambda \leq 1$.

average MAD estimate (averaged across the two choice sets) was 3%. On the basis of this estimate and Equation 11, we calculated the improvement in predictive validity obtained by using each of the alternative models rather than the standard VMM and the naive GCM. As Table 10 shows, across the two choice sets, all five competing models provided substantial improvements (48%–68%) over the VMM. Furthermore, the contextual concavity models and the LAM provided large improvements (26%–38%) over the GCM, which demonstrates that the leading models yield improved aggregate predictive validity by accounting for context effects rather than just correcting for possible biases in utility measurement (stated preferences versus revealed preferences).

Individual-level predictive validity. We estimated the models' parameters using one choice-set condition, and using the conjoint partworths of a participant in the other choice-set condition, we predicted the choice probabilities in that latter set (we applied the same procedure to the other choice set). It is noteworthy that unlike the first empirical application, we separated the partworths elicitation from the preceding choice by a (ten-minute) filler task, which reduces the possibility that choice affected the estimated partworths. Table 10 shows that, pooled across the two choice sets, the individual-level measure of predictive validity favors the two contextual concavity models and the LAM over the VMM, GCM, and RAM.

Model fit. As is shown in Table 10, the three leading models yielded superior BIC measures than the VMM, GCM, and RAM (in all nine comparisons). In addition, the likelihood ratio test indicates that all the models provided a significant improvement over the VMM, though the improvements were most pronounced for the three leading context-dependent models.

Additional evidence. The various measures of predictive validity and fit provide strong support for the proposition that accounting for the local choice context can improve the performance of choice models, beyond any improvements that arise from adjustments to utility measurement. Indeed, although the GCM outperformed the simpler VMM, both were far inferior to the contextual concavity models and the LAM.

As is shown in Table 8, the context-dependent models were also better able to predict the strong compromise effects that we observed in the choice data; this was particularly true for the contextual concavity models. In contrast, the context-independent models did not predict the compromise effects. Moreover, consistent with betweenness inequality, the VMM even predicted reversed (negative) compromise effects for all five possible measures. Accordingly, the average compromise measure MAD (calculated across the five measures) between the predicted and observed compromise effects was the best for the CCM and NCCM (4% and 3%, respectively) and the worst for the VMM and GCM (22% and 21%, respectively).

GENERAL DISCUSSION

The compromise effect is a major finding documented in behavioral marketing and decision research. It has substan-

Table 10

COMPARISON OF MODELS' FIT AND PREDICTIVE VALIDITY (ACROSS CHOICE SETS IN EMPIRICAL APPLICATION 2)

	VMM	*GCM*	*CCM*	*NCCM*	*LAM*	*RAM*
Portable PCs						
Aggregate-level prediction (MAD)	9.6%	6.4%	5.1%	5.5%	5.3%	6.0%
Improvement in aggregate-level prediction (MAD) over VMM (Equation 11)	—	48%	68%	62%	65%	54%
Improvement in aggregate-level prediction (MAD) over GCM	—	—	38%	26%	32%	12%
Individual-level prediction (–2LL)	565.6	561.8	534.4	535.0	532.1	575.1
Model fit (BIC)	570.3	557.9	544.1	541.8	538.4	571.2
Number of parameters	1	2	2	2	2	4

tial implications for consumer choice, and it represents a significant violation of standard microeconomic theory. Nevertheless, so far, there have been no attempts to incorporate the compromise effect in formal choice models and to test the fit and predictive validity of such models. Indeed, in an address to the 2001 University of California–Berkeley Choice Symposium, the Nobel laureate Daniel McFadden called for the incorporation of behavioral context effects in choice models and for a comparison of such models with existing practice to establish the contribution. We take a step in this direction by incorporating the compromise effect in four context-dependent choice models and by comparing the alternative models with one another, with the standard VMM, and with a stronger naive model that adjusts for possible biases in utility measurement. As we discuss subsequently, our models can theoretically capture other context effects, including the well-known "asymmetric dominance effect" (Huber, Payne, and Puto 1982).

The current research can be viewed as part of the ongoing (fruitful) attempt to bridge the consumer behavior and marketing science disciplines. Such cross-fertilization often involves choice modeling based on theory and empirical generalizations from decision research, social psychology, and consumer behavior (e.g., Bell and Lattin 2000; Hardie, Johnson, and Fader 1993; Winer 1986). It should be emphasized that the contribution of such research is maximized when choice models are actually calibrated and validated using empirical data, not just postulated at an axiomatic level.

Main Findings and Theoretical Implications

Using two empirical applications, we estimated and then tested four alternative choice models that incorporate the compromise effect. The two empirical applications used different preference-elicitation methods (self-explicated approach versus full-profile conjoint analysis) to estimate the (context-independent) partworths. Furthermore, whereas the first empirical application tested the alternative models using the traditional design of the compromise effect (i.e., three options defined on two attributes), the second application employed a more complex choice setting with five alternatives defined on four attributes. The use of different utility-measurement techniques and increased dimensionality allows for generalization of the results that pertain to the compromise effect and the calibration and validation of the models. To the best of our knowledge, this article is the first to demonstrate robust, systematic compromise effects in choices that involve more than three options and two product attributes. However, it is noteworthy that the average compromise effect was greater in the first empirical application (24%) than in the second application (17%). A possibility that merits further research is that increasing the dimensionality and/or choice-set size attenuates the compromise effect.

A main result is that accounting for the local choice context or the relative positions of options in the choice set can significantly improve predictive validity and fit over the standard choice model. However, an alternative explanation for the superiority of such models to the VMM is that they adjust for possible discrepancies between the measured (partworth) utilities and the revealed utilities (based on choices). To rule out this rival account, we tested a stronger context-independent model: the GCM. The GCM is similar to the contextual concavity models because it adds a layer of concavity to the estimated partworths and can correct the measured utilities to be consistent with choices. Yet unlike the contextual concavity models, in the GCM, the induced concavity and possible adjustment of utilities is done at a global level, independent of the local choice context.

The validation and fit measures indicate that the three leading models outperformed the VMM and the GCM. In particular, after we accounted for the sampling-based error, the aggregate-level cross–choice set predictions (which is probably the most relevant prediction criterion from a managerial point of view) showed that the contextual concavity models and the LAM yielded major improvements in the MAD measure over the VMM and the GCM (the improvements ranged from 46% to 68% and from 26% to 50%, respectively). The NCCM, CCM, and LAM also predicted the enhanced shares of the compromise options and the observed compromise effects. In contrast, the GCM and RAM were unable to predict the compromise effects, and the VMM (and at times the GCM) even predicted reversed (negative) compromise effects. These findings provide strong support for the notion that the improved performance of the models was driven, at least in part, by their context-dependent component.

A limitation of the empirical applications is that we did not calibrate the model parameters (e.g., the concavity and loss-aversion parameters) at the individual level (though we fully accounted for heterogeneity in the utility functions). It is likely that because of individual differences in the decision process and susceptibility to the compromise effect, consumers differ with regard to their context-dependent effect. Therefore, further research can use methods such as hierarchical Bayes (e.g., Rossi and Allenby 2003) to capture heterogeneity in the context-effect parameters. Next, we compare the alternative models in terms of their conceptual characteristics and ability to account for additional context effects.

Conceptual Similarities and Differences Among the Alternative Models

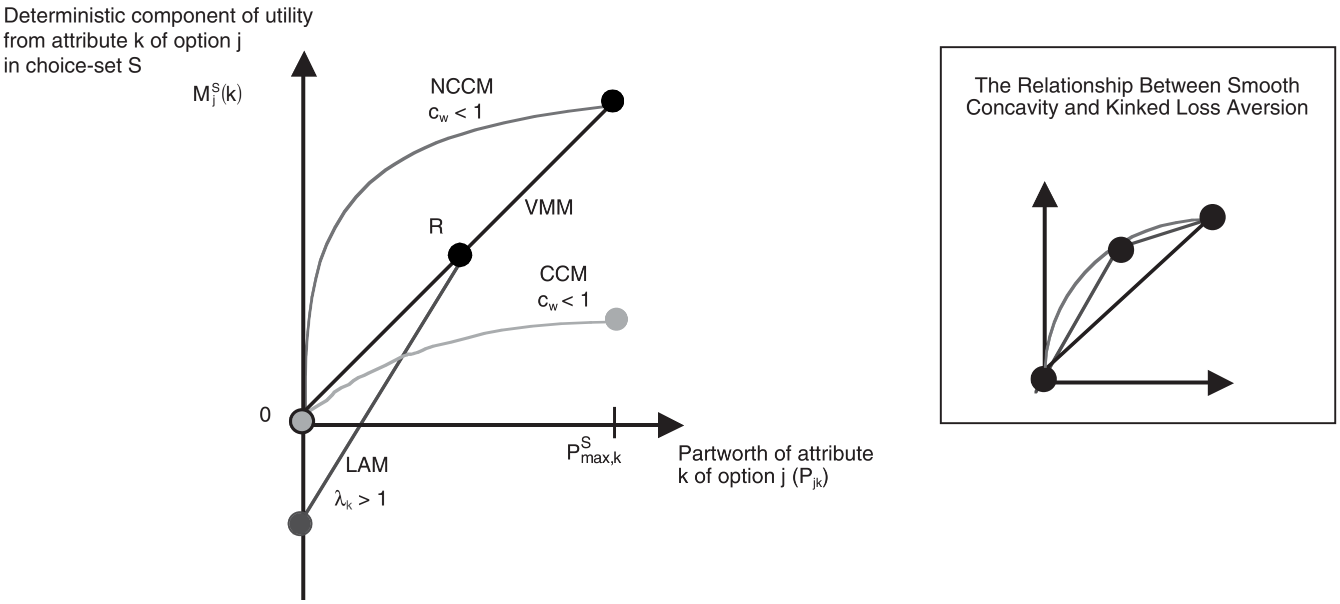
All four alternative models operate as if consumers have some absolute (context-independent) utilities (i.e., valuations for different attribute levels), but the utilities are affected by the relative positions of the choice options (i.e., by the local choice context). Thus, the models depict consumer choice as context and reference dependent. Next, we elaborate on some of the distinctions underlying the alternative models.

Effects on attribute importance weights. Figure 6 illustrates the relationship between the three leading models and the VMM.[17] Figure 6 highlights that only the NCCM retains the original attribute importance weights, as defined by the range of the partworths in the local choice set. Both the CCM and the LAM distort the attribute importances; in the CCM, greater concavity (smaller value of c_k) decreases

[17]To place the models on a comparable scale, we rescaled the context-independent partworths by subtracting $P^S_{min,k}$ from each partworth. We also rescaled the LAM by adding P^S_{Rk} to each partworth.

Figure 6

A GRAPHICAL ILLUSTRATION OF THE RELATIONSHIP AMONG THE ALTERNATIVE MODELS



the attribute importance, whereas in the LAM, greater loss aversion enhances the attribute importance. It is noteworthy that because the NCCM preserves the original importance weights, this model rules out an alternative explanation that the enhanced performance of the models is due to differences between stated and revealed preferences in terms of attribute importances (see Heath et al. 2000).

The relationship between concavity and loss aversion. Figure 6 also illustrates that a smooth (continuous) concave function and a kinked loss-aversion function are good approximations for each other. Indeed, it is interesting that diminishing returns, which underlie the nonkinked contextual concavity models, imply that the impact of a loss on an attribute always outweighs the impact of a corresponding gain (with any point on the function as a reference point). Thus, diminishing sensitivity to gains (i.e., concavity) is similar to increased sensitivity to losses (i.e., loss aversion). In light of this similarity, it is not surprising that the two models perform similarly in our empirical applications. However, whereas the LAM employs constant loss aversion, the contextual concavity models imply increasing loss aversion.

Single versus multiple reference points. Whereas all the alternative models are reference dependent, the contextual concavity models and the LAM employ a single reference point (i.e., the partworth of the lowest and midrange attribute levels, respectively), and the RAM uses multiple references. More specifically, the RAM assumes that each alternative is evaluated against all of the other options in the choice set, in what could be described as a tournament (see Tversky and Simonson 1993, p. 1185). We also tested an additional family of tournament models, the multireference LAM, in which each option is evaluated on the basis of its gains and losses compared with those of all other options in the choice set.[18] On both modeling and behavioral grounds, we find tournament models such as the multireference LAM and the RAM less appealing than unireference models. From a modeling perspective, in Pareto-optimal choice sets, tournament models lead to estimation problems due to multicollinearity between the magnitudes of the multiple advantages (gains) and disadvantages (losses). This multicollinearity may have contributed to the underperformance of the RAM. From a behavioral standpoint, tournament models are less feasible than single reference models because they place inordinate processing demands on consumers (see, e.g., Shugan 1980).

Parsimony in the parameter space. The alternative models also differ with regard to their number of parameters. In particular, the contextual concavity models and the LAM not only provide better predictive validity and fit but also require estimation of fewer parameters (d + 1 parameters for the contextual concavity models and the LAM versus 2d + 2 parameters for the RAM, where d is the number of product dimensions). Furthermore, the results in Empirical Application 1 indicated that the concavity parameter of the NCCM did not significantly vary across the two product

[18]The most general form of the multireference LAM can be written as follows:

$$M_j^S = \sum_{h \neq j \in S} \Big[\sum_k g_k \times |P_{jk} - P_{hk}| \times 1(P_{jk} > P_{hk}) - \sum_k l_k \times |P_{jk} - P_{hk}| \times 1(P_{jk} < P_{hk})\Big],$$

where $g_k[l_k]$ is the relative gain (loss) parameter for attribute k. Although the multireference LAMs had acceptable predictive validity, their gain parameters were counterintuitively negative because of the models' substantial gain–loss multicollinearity.

attributes in either product category; thus, the NCCM requires estimation of the lowest number of parameters (i.e., two). The excellent performance of the NCCM with a uniform concavity parameter in both empirical applications represents an advantage for the NCCM over the other alternative models.

Our empirical results do not provide a clear choice among the three winning models, namely, the CCM, the NCCM, and the LAM. Further research is required to compare the alternative models in settings different from the one we employed. For example, as we discuss next, the models differ in their ability to capture other choice-set (context) effects, which provides a fertile ground for future tests.

Modeling other context effects. Several behavioral context effects other than the compromise effect and extremeness aversion have been documented, including asymmetric dominance, asymmetric advantage, enhancement, and detraction (for details, see Huber, Payne, and Puto 1982; Huber and Puto 1983; Simonson and Tversky 1992). Under most conditions, the contextual concavity models and the LAM can conceptually capture the asymmetric dominance and advantage effects. However, in certain specific cases (when the added alternative does not affect the attribute ranges), unireference models such as the contextual concavity models and the LAM cannot capture these context effects, whereas tournament (multireference) models can. In addition, although the contextual concavity models and the LAM effectively incorporate the enhancement effect, only the LAM can simultaneously capture detraction and compromise.[19]

In conclusion, the alternative models can theoretically account for a wide range of context effects, including compromise, polarization, asymmetric dominance, and other local contrast effects. Another model that can theoretically capture multiple context effects is the loss-aversion–based centroid model of Bodner and Prelec (1994). Further research can construct and empirically test a unifying model that accounts for the greatest number of context effects across the widest range of choice situations.

Practical Implications

Beyond the theoretical significance of incorporating the compromise effect in formal choice models, this issue has important practical implications. First, consider conjoint choice simulators, which define a market scenario and a set of competitive products, and then perform sensitivity analyses. Such simulators employ the standard value-maximization utility function, which is estimated with one of various traditional methods that are not necessarily choice based, such as full-profile, trade-off tables, adaptive conjoint analysis, self-explicated, and hybrid (see Green and Srinivasan 1990). However, as the present research demonstrates, by neglecting context effects, the VMM will lead to inaccurate estimates of the choice probabilities. Thus, by using the proposed models, choice simulations can potentially yield more accurate predictions and analyses. It would also be worthwhile to test the alternative models using utilities obtained from adaptive conjoint analysis, full-profile ratings, and other common preference-measurement techniques.

[19]Further details are provided in the work of Kivetz, Netzer, and Srinivasan (2004).

Second, in addition to possible improvements in the prediction of consumer choice, managers can use a context-dependent model such as the NCCM to define the optimal set of product or service offerings so as to maximize product-line profitability. That is, the alternative models enable more accurate calculation of the predicted choice shares for any given product-line portfolio. Furthermore, the pervasiveness of product offerings in the marketplace assorted according to the underlying structure of the compromise effect (see Figure 2 and previous discussion) suggests that marketers can employ context-dependent models to construct product menus that enhance the share of high-margin products.

Recent statistical advances that enable the obtaining of individual-level estimates from scarce data (i.e., hierarchical Bayes models) have promoted the application of methods that use choices to estimate utility functions, such as choice-based conjoint (CBC). A notable question arises about the applicability of the models proposed herein to such domains in which the partworths are estimated directly from choices (and simultaneously with the model parameters). On the one hand, CBC may minimize context effects during model estimation because of the use of within-subjects repeated choices (see, e.g., Huber, Payne, and Puto 1982; Kahneman and Tversky 1996), thus reducing the usefulness of our models. On the other hand, predictions derived from CBC may sometime be inaccurate, because the compromise effect is likely when consumers make a real purchase decision, as in Figure 2.

Finally, it is important to note that the present results will not apply to market choices that obscure the identity of compromise options. For example, many purchase decisions involve cases in which certain attribute values are missing for some of the products (e.g., Kivetz and Simonson 2000). Furthermore, the compromise effect may be less likely in CBC-type choice sets chosen from experimental designs that are not located on (or near) the efficient frontier or that comprise many attributes that cannot be represented with a one-dimensional subspace of options (e.g., price versus overall quality trade-offs in Figure 2, memory and hard drive versus speed and battery life in Table 7). In addition, consumers may face many alternatives at the point of purchase, but they may choose from a narrower consideration set that is unobservable to the researcher (as is often the case in scanner-panel data). Indeed, searching for and modeling the compromise effect and, more generally, other behavioral context effects using marketplace data and marketing science applications is fertile ground for further research.

REFERENCES

Assar, Amardeep and Dipankar Chakravarti (1984), "Attribute Range Knowledge: Effects on Consumers' Evaluation of Brand-Attribute Information and Search Patterns in Choice," in *AMA Educators' Proceedings*, Russell W. Belk et al., eds. Chicago: American Marketing Association, 62–67.

Bell, David R. and James M. Lattin (2000), "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity," *Marketing Science*, 19 (2), 185–200.

Benartzi, Shlomo and Richard Thaler (2002), "How Much Is Investor Autonomy Worth?" *Journal of Finance*, 57 (4), 1593–1616.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (December), 187–217.

Bodner, Ronit and Drazen Prelec (1994), "The Centroid Model of Context-Dependent Choice," working paper, Sloan School of Management, Massachusetts Institute of Technology.

Bretton-Clark (1989), *Conjoint Linmap*. Bernardsville, NJ: Bretton-Clark.

Chernev, Alexander (2004), "Extremeness Aversion and Attribute Balance Effects in Choice," *Journal of Consumer Research*, 31 (September), forthcoming.

Dhar, Ravi, Stephen M. Nowlis, and Steven J. Sherman (2000), "Trying Hard or Hardly Trying: An Analysis of Context Effects in Choice," *Journal of Consumer Psychology*, 9 (4), 189–200.

——— and Itamar Simonson (2003), "The Effect of Forced Choice on Choice," *Journal of Marketing Research*, 40 (May), 146–60.

——— and Klaus Wertenbroch (2000), "Consumer Choice Between Hedonic and Utilitarian Goods," *Journal of Marketing Research*, 37 (February), 60–71.

Drolet, Aimee L. (2002), "Inherent Rule Variability in Consumer Choice: Changing Rules for Change's Sake," *Journal of Consumer Research*, 29 (December), 293–305.

———, Itamar Simonson, and Amos Tversky (2000), "Indifference Curves That Travel with the Choice Set," *Marketing Letters*, 11 (3), 199–209.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research," *Journal of Consumer Research*, 5 (September), 103–123.

——— and ——— (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54 (October), 3–19.

Hall, Bronwyn H. and Clint Cummins (1999), *Time Series Processor Version 4.5 Reference Manual*. Palo Alto, CA: TSP International.

Hardie, Bruce, Eric J. Johnson, and Peter Fader (1993), "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," *Marketing Science*, 12 (4), 378–94.

Heath, Chip, Richard P. Larrick, and George Wu (1999), "Goals as Reference Points," *Cognitive Psychology*, 38 (February), 79–109.

Heath, Timothy B., Gangseog Ryu, Subimal Chatterjee, Michael S. Mccarthy, David L. Mothersbaugh, Sandra Milberg, and Gary J. Gaeth (2000), "Asymmetric Competition in Choice and the Leveraging of Competitive Disadvantages," *Journal of Consumer Research*, 27 (December), 291–308.

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

——— and Christopher P. Puto (1983), "Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects," *Journal of Consumer Research*, 10 (June), 31–44.

———, Dick R. Wittink, John A. Fiedler, and Richard Miller (1993), "The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice," *Journal of Marketing Research*, 30 (February), 105–114.

Hutchinson, J. Wesley, Wagner A. Kamakura, and John G. Lynch (2000), "Unobserved Heterogeneity as an Alternative Explanation for 'Reversal' Effects in Behavioral Research," *Journal of Consumer Research*, 27 (December), 324–44.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47 (2), 263–91.

——— and ——— (1996), "On the Reality of Cognitive Illusions," *Psychological Review*, 103 (3), 582–91.

Kalyanaram, Gurumurthy and Russel S. Winer (1995), "Empirical Generalizations from Reference Price and Asymmetric Price Response Research," *Marketing Science*, 14 (3), 161–69.

Kivetz, Ran (2003), "The Effects of Effort and Intrinsic Motivation on Risky Choice," *Marketing Science*, 22 (4), 477–502.

———, Oded Netzer, and V. Srinivasan (2004), "Extending Compromise Effect Models to Complex Buying Situations and Other Context Effects," *Journal of Marketing Research*, 41 (August), 262–68.

——— and Itamar Simonson (2000), "The Effects of Incomplete Information on Consumer Choice," *Journal of Marketing Research*, 37 (November), 427–48.

Lehmann, Donald R. and Yigang Pan (1994), "Context Effects, New Brand Entry, and Consideration Sets," *Journal of Marketing Research*, 31 (August) 364–74.

Leigh, Thomas W., David B. MacKay, and John O. Summers (1984), "Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison," *Journal of Marketing Research*, 21 (November), 456–62.

McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, Paul Zarembka, ed. New York: Academic Press, 105–142.

Meyer, Robert J. and Eric J. Johnson (1995), "Empirical Generalizations in the Modeling of Consumer Choice," *Marketing Science*, 14 (3), 180–89.

Nowlis, Stephen M. and Itamar Simonson (2000), "Sales Promotions and Choice Context as Competing Influences on Consumer Decision Making," *Journal of Consumer Psychology*, 9 (1), 1–17.

Payne, John W., James R. Bettman, and Eric J. Johnson (1992), "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, 43, 87–131.

Prelec, Drazen, Birger Wernerfelt, and Florian Zettelmeyer (1997), "The Role of Inference in Context Effects: Inferring What You Want from What Is Available," *Journal of Consumer Research*, 24 (June), 118–25.

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304–328.

Shugan, Steven M. (1980), "The Cost of Thinking," *Journal of Consumer Research*, 7 (2), 99–111.

Simonson, Itamar (1989), "Choice Based on Reasons: The Case of Attraction and Compromise Effects," *Journal of Consumer Research*, 16 (September), 158–74.

——— and Stephen M. Nowlis (2000), "The Role of Explanations and Need for Uniqueness in Consumer Decision Making: Unconventional Choices Based on Reasons," *Journal of Consumer Research*, 27 (June), 49–68.

——— and Amos Tversky (1992), "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29 (August), 281–95.

Srinivasan, V. (1988), "A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences," *Decision Sciences*, 19 (Spring), 295–305.

——— and Chan Su Park (1997), "Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement," *Journal of Marketing Research*, 34 (May), 286–91.

Thaler, Richard H. (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (3), 199–214.

Tversky, Amos and Daniel Kahneman (1991), "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106 (November), 1039–61.

——— and Itamar Simonson (1993), "Context-Dependent Preferences," *Management Science*, 39 (10), 1179–89.

Viscusi, W. Kip, Wesley A. Magat, and Joel Huber (1987), "An Investigation of the Rationality of Consumer Valuations of Multiple Health Risks," *RAND Journal of Economics*, 18 (4), 465–79.

Wernerfelt, Birger (1995), "A Rational Reconstruction of the Compromise Effect: Using Market Data to Infer Utilities," *Journal of Consumer Research*, 21 (March), 627–33.

Winer, Russell S. (1986), "A Reference Price Model of Demand for Frequently-Purchased Products," *Journal of Consumer Research*, 13 (September), 250–56.

Wright, Peter (1975), "Consumer Choice Strategies: Simplifying Versus Optimizing," *Journal of Marketing Research*, 12 (February), 60–67.