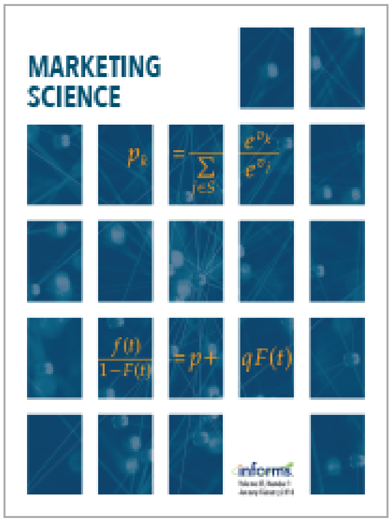

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test

Linda Court Salisbury, Fred M. Feinberg,

MARKETING SCIENCE
Vol. 29, No. 1, January–February 2010, pp. 1–17
ISSN 0732-2399 | EISSN 1526-548X | 10 | 2901 | 0001

DOI 10.1287/mksc.1080.0464

Alleviating the Constant Stochastic Variance Assumption in Decision Research: Theory, Measurement, and Experimental Test

Linda Court Salisbury
Carroll School of Management, Boston College, Chestnut Hill, Massachusetts 02467,
salisbli@bc.edu

Fred M. Feinberg
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109,
feinf@umich.edu

Analysts often rely on methods that presume constant stochastic variance, even though its degree can differ markedly across experimental and field settings. This reliance can lead to misestimation of effect sizes or unjustified theoretical or behavioral inferences. Classic utility-based discrete-choice theory makes sharp, testable predictions about how observed choice patterns should change when stochastic variance differs across items, brands, or conditions. We derive and examine the implications of assuming constant stochastic variance for choices made under different conditions or at different times, in particular, whether substantive effects can arise purely as artifacts. These implications are tested via an experiment designed to isolate the effects of stochastic variation in choice behavior. Results strongly suggest that the stochastic component should be carefully modeled to differ across both available brands and temporal conditions, and that its variance may be relatively greater for choices made for the future. The experimental design controls for several alternative mechanisms (e.g., flexibility seeking), and a series of related models suggest that several econometrically detectable explanations like correlated error, state dependence, and variety seeking add no explanatory power. A series of simulations argues for appropriate flexibility in discrete-choice specification when attempting to detect temporal stochastic inflation effects.

Key words: brand choice; choice models; decisions under uncertainty; decision making over time; econometric models; lab experiments; measurement and inference; probability models; simulation; stochastic models
History: Received: March 27, 2008; accepted: September 10, 2008; processed by Gary Russell. Published online in *Articles in Advance* November 3, 2009.

Introduction

Research in marketing focuses, by and large, on theories of behavior expressed as concrete hypotheses. Verifying such theories overwhelmingly involves experimental manipulations or examining field data for differences in mean, expected, or average responses, even at the individual level. By contrast, stochasticity—whether it is conceptualized as noise, uncertainty, response variability, or in some other way—is typically treated (if considered at all) as residing in the denominator: something to minimize to make mean effects stand out, unmodeled "error" in some ancillary statistical model. Such methods can presume, often tacitly, that differences in behavioral outcomes (e.g., observed choices) are driven primarily by (differences in) what is predictable and deterministic, not what is unobserved and stochastic. Recent research, however, has cautioned researchers to be mindful that different conditions can entail differences in response variability as well (Louviere 2001, Louviere et al. 2002). Employing analysis methods—notably, choice models—that do not allow for (or estimate) differences in response variation can introduce potential misspecifications, ones with important substantive implications.

One excellent example, widely studied by behavioral researchers, involves temporal effects on behavior. Most consumer decisions are not only made over time, but also involve outcomes—products, services, investments, etc.—experienced or consumed at some point in the future. That is, consumers often make sequences of choices, each intended to anticipate that consumer's preferences at some future consumption time. Given the ubiquity of this scenario, a great deal of research in both marketing and behavioral decision theory has attempted to develop, test, and measure key aspects of theories of how choices are made, both dynamically and prospectively. For example, it has been established that consumers making multiple choices tend to opt for greater variety when all items are chosen now for future consumption, rather than later at the time of consumption (Read and

Loewenstein 1995, Simonson 1990, Walsh 1995); that decision makers are more likely to prefer hedonic indulgences over cash of equal or greater value when there is a time delay between the decision and its outcome (Kivetz and Simonson 2002); and that people are more likely to choose a "virtue" rather than a "vice" for future consumption, whereas the opposite holds for immediate consumption (Read et al. 1999).

Understanding temporal aspects of choice requires the analyst to account for *uncertainty* in verifying any posited effects, such as decision avoidance (Anderson 2003, Dhar 1997); preference for more immediate, smaller gains (Keren and Roelofsma 1995); and diversification (Simonson 1990), among many others. As noted at the outset, such effects are typically studied as a component of the deterministic, as opposed to the stochastic, portion of utility. It is customary in the psychological and statistical literatures to treat "error" as if it were amorphous, a by-definition non-explicable construct that lacks meaningful patterns or structure. However, in recent years, researchers have challenged this view of "error," including its terminology, referring instead to "unobserved variability." This is consistent with a critical distinction in hierarchical models, where the higher-level specifications distinguish "observed" and "unobserved" heterogeneity. In this article, although we hew to the simpler and more standard term "error," a core concept is that put forward by Louviere et al. (2002), who sought to "dissect" the random or "unobserved" component of utility and who suggested numerous dimensions across which the variance of this unobserved utility could vary. Here, we focus on verifying and measuring that for a specific source, involving choices made prospectively versus at the time of consumption. Beyond helping to understand behavioral drivers, this serves to free random utility models from misidentification based on assuming utility variance equal across brands and temporal conditions. Such misspecifications can potentially alter whether a given behavioral hypothesis is supported or refuted by the data, and it is this we seek to examine.

Specifically, we examine whether substantive effects can arise not only from deterministic shifts in brand evaluation, but from swells in unobserved stochastic variation of the sort disallowed by commonly used discrete-choice specifications and other regression-based methods (Adamowicz et al. 2008). That is, certain theoretical conclusions may be at least in part statistical artifacts, if one does not apply equal care to specifying the deterministic and stochastic nature of utility and preference. Although this issue may appear at first blush to be of mainly technical interest, it goes to the heart of what drives differences in observed choice patterns: do they primarily result from deterministic influences (perhaps suggesting biased decision processes, inattention, or misunderstanding), from increased environmental uncertainty, or from some amalgam? Specifically, when modeling choice, is there a compelling reason to allow for a flexible, perhaps even parametric, account of the stochastic term?

Prior literature in econometrics has amply addressed various sorts of nonconstant error variance. For example, the chief motivation behind the GARCH class of models (Engle 1982) is to account for heteroscedasticity, and such models are cornerstones in finance and econometrics in general. Swait and Louviere (1993) argue persuasively that when modeling multiple "sources" of choice data, estimation of substantively important quantities is confounded with error variance, and that error "scale ratio" parameters should be estimated as well. In their landmark treatment of the econometric modeling of choice data, Louviere et al. (2000) devote many chapters to various stratagems for alleviating the i.i.d. error assumption, among them error scaling, mixed logit, and general probit models. A key application of error scaling of the type we will explore here combines sources of "revealed" and "stated" preference data (e.g., Hensher et al. 1999) by setting various parameters equal across data sets and allowing error variances to differ. Even in stated preference experiments, systematic differences in stochastic variation have been found to accompany effects of rank order and fatigue (Bradley and Daly 1994).

Our goal is to investigate the effects of presuming a constant value for unobserved stochastic variation across items and over time. To examine this issue, we take a multitiered approach, involving derivation, model formulation, an experiment, parameter estimation, and simulation. First, we demonstrate that firm directional predictions can be made about which items' choice probabilities will, and will not, be affected by "stochastic inflation." This allows the development of a formal model of the effects of variation in the unobserved stochastic component of utility; the model generates testable predictions about which items should display shifts in choice probability (namely, one's most- and least-favored items), and affords the use of discrete-choice-based methods to analyze participants' choice patterns. We then show how the total error common in the random utility framework can be decomposed into that for the analyst and that for the consumer, both of whom must "predict" *experienced* utility; and, moreover, that a firm lower bound on the consumer's (unobservable) "temporal stochastic inflation factor" can in fact be extracted from quantities measurable via discrete-choice methods. Next, we present an experiment designed to verify whether specific choice behavior patterns can stem from increased stochastic variation, and to explore the marginal explanatory power of both the posited, temporally induced effects and several others suggested by prior literature and random utility theory. This allows a sequence of econometric models to test for and rule out a number of alternative effects, while providing strong support for not restricting stochastic variance to be constant across temporal conditions (as well as more modest support for an analogous statement across brands/items). Finally, we demonstrate via simulation that one could draw erroneous conclusions about purported (deterministic) effects when using traditional statistical methods that may be misspecified.

Time Delay, Uncertainty, and Choice Probability Distribution

Throughout our discussion, we will compare "future" and "immediate" decision modes. The former will connote making choices *now* for *future* consumption; the latter will mean that choices are made at the time of consumption. In our experiment and the tests that follow, we will further specify, and compare an example of, these two types of decision mode.

It is well known that decision makers have difficulty estimating future utility (Kahneman and Snell 1992, Kahneman et al. 1997), and whether choosing for future or immediate consumption, decision makers experience uncertainty to some degree. Preference uncertainty arises when one cannot predict perfectly which of the available choice alternatives will maximize *experienced* utility. It can be influenced by any of a number of contextual variables, such as the number of available alternatives (Iyengar and Lepper 2000) or their similarity (Dhar 1997). We focus on the effect of time delay between choosing and consuming, specifically in its leading to greater preference uncertainty (Simonson 1990). Note that this phenomenon is distinct from the primary focus of the literature on intertemporal choice (e.g., choosing between two alternatives whose outcomes are experienced at different times).

One need not presume that the *expected* utility of available alternatives is affected by choice mode to account for between-mode differences. Suppose, for example, that one is choosing between two alternatives, A and B, and alternative B has greater mean utility than alternative A ($V_B > V_A$).[1] However, if variance is greater when choosing for future rather than immediate consumption, the probability that the experienced utility of A will exceed the experienced utility of B is also greater; consequently, decision makers would be more likely to choose alternative A in the future than in the immediate consumption mode. This assertion is easily verified in a random utility framework, in a manner to be generalized shortly.

Let us introduce a relative error-variance scaling parameter, $\sigma \geq 1$ (Swait and Louviere 1993),

$$U_{At} = V_{At} + \sigma \varepsilon_{At}, \tag{1}$$

where, without loss of generality, the scaling factor is normalized to $\sigma = 1$ for immediate consumption ($t = 1$), so $\sigma > 1$ for choices made for future consumption ($t > 1$). The probability of choosing alternative A (with the t subscript dropped for simplicity) is

$$\begin{aligned} P_A &= \Pr(V_A + \sigma\varepsilon_A > V_B + \sigma\varepsilon_B) \\ &= \Pr\left(\varepsilon_B - \varepsilon_A < \frac{V_A - V_B}{\sigma}\right). \end{aligned} \tag{2}$$

We can consider the difference of the errors, $\varepsilon_B - \varepsilon_A$, a random variable in its own right, and examine expression (2) as a function of σ.[2] Because the mean anticipated utility of alternative B is greater than that for A, $(V_A - V_B)$ is negative, and so $(V_A - V_B)/\sigma$ is increasing in σ. Thus, the probability of choosing the alternative with the smaller mean anticipated utility (alternative A) monotonically increases (towards 1/2) in σ, whereas that of the other alternative (B) decreases.

Of course, real decisions typically involve more than two potential options. We show in the appendix that in a general random utility setting, one can make clear directional predictions about which items' probabilities change with the error-scaling factor, σ, when there are multiple options from which to choose. Specifically, we prove three facts: when degree of stochasticity (σ) increases, (1) the most-favored (highest v_j) item is less likely to be chosen, (2) the least-favored (smallest v_j) item is more likely to be chosen, and (3) systematic predictions cannot be made about other items. It is remarkable that the same is *never* true for the most- and least-favored alternatives, regardless of the joint error density, and so theory-driven consequences of increasing uncertainty will concern only individual-specific "favorites" and "least favorites."

In summary, we expect that, ceteris paribus, choices made for future consumption will entail a greater degree of stochastic variation than those made for immediate consumption. Note that this is not the case

[1] We distinguish this from "expected" utility because this connotes the standard expectation operator, and there is no guarantee that the error process is zero mean (e.g., it is not for Gumbel errors, which underlie logit models). Thus, "mean anticipated utility" is associated with the "deterministic portion of utility."

[2] Note that the error terms, ε_A and ε_B, need not be i.i.d. but may have an arbitrary nondegenerate bivariate distribution; nor need they have zero mean (as with Gumbel distributions). However, although not necessary here, it would not be unreasonable to presume that ε_A and ε_B have the same form of *marginal* distribution, an issue we take up later in the empirical application.

for *specific* times in the future; for example, there is more uncertainty about a typical American's dinner preference for turkey tonight versus for next Thanksgiving Day. We emphasize that random utility (RU) theory makes testable directional predictions—when choices are made for future consumption, the probability of choosing the most-favored item decreases and the probability of choosing the least-favored item increases, relative to choices for immediate consumption—without requiring systematic changes to the deterministic component of utility, if consumption time delay entails relatively greater stochastic variation. We next examine conditions under which statements can be made about the consumer's (unobservable) temporal inflation factor, based on measurements stemming from discrete-choice models, which also include error for predictions made by the analyst.

Consumer, Analyst, and Total Error

Throughout, we appeal to discrete-choice models based on the standard random utility framework that underlies the majority of empirical work in marketing and behavioral research; unless needed for clarity, we suppress subscripts for option (brand), time, and consumer. The analyst can only imperfectly estimate the consumer's assessment of his or her utility, and so includes error process ε_T; this is the (total) error typically specified in RU models, where all error is presumed to stem from analysts' inability to perfectly model consumer utility *at the time of choice*. As suggested explicitly by Louviere (2001, p. 507, Equation (4)), this total error can be decomposed into subcomponents "that reference within-subjects, between-subjects, between-contexts, between-measurement instruments, between-time periods, and so forth components of response variability." Here, because consumption is not necessarily immediate—consumers need to anticipate their *experienced* utility at the time of (future) *consumption* (Kahneman et al. 1997)—we decompose ε_T into subcomponents ε_A and ε_C to represent analyst error and consumer error, respectively (we will later account parametrically for differences in response variability between brands and between temporal conditions). These error processes have some joint distribution $(\varepsilon_A, \varepsilon_C)$, about which we make no assumptions. Total error, reflected in the measured system and all observations, is given by $\varepsilon_T = \varepsilon_A + \varepsilon_C$, and so we may write

$$\begin{aligned} U &= V + \varepsilon_T \\ \varepsilon_T &= \varepsilon_A + \varepsilon_C \\ \sigma_T^2 &= \sigma_A^2 + \sigma_C^2 + 2\,\mathrm{Cov}(\varepsilon_A, \varepsilon_C). \end{aligned} \tag{3}$$

Measuring the ratio in error variances between the future (F) and immediate (I) choice conditions is only possible for the *total* error, $\sigma^2_{F,T}/\sigma^2_{I,T}$. However, we are concerned throughout with the ratio of the *consumer's* error variances, $\sigma^2_{F,C}/\sigma^2_{I,C}$, and so proceed as follows. Given that analyst error is unobservable, we invoke the standard model identification assumption that $\sigma^2_{I,A} = \sigma^2_{F,A}$ (this will be relaxed later). For simplicity of derivation, let $a = \sigma_{I,A} = \sigma_{F,A}$, $r = \sigma_{F,T}/\sigma_{I,T}$, and $\rho = \mathrm{Corr}(\varepsilon_A, \varepsilon_C)$, in which case we can write

$$\begin{aligned} \frac{\sigma^2_{F,T}}{\sigma^2_{I,T}} &= \frac{\sigma^2_{F,A} + \sigma^2_{F,C} + 2\rho\sigma_{F,A}\sigma_{F,C}}{\sigma^2_{I,A} + \sigma^2_{I,C} + 2\rho\sigma_{I,A}\sigma_{I,C}} \\ \Rightarrow\ r^2 &= \frac{a^2 + \sigma^2_{F,C} + 2\rho a\sigma_{F,C}}{a^2 + \sigma^2_{I,C} + 2\rho a\sigma_{I,C}} \\ &= \frac{a^2(1-\rho^2) + (\sigma_{F,C} + \rho a)^2}{a^2(1-\rho^2) + (\sigma_{I,C} + \rho a)^2}. \end{aligned} \tag{4}$$

To establish a relationship between the error variance ratio for the consumer, $\sigma_{F,C}/\sigma_{I,C}$, and that in total, r, we note that removing the same positive quantity from the numerator and denominator of a ratio greater than one will increase its value; thus,

$$\begin{aligned} r^2 &= \frac{a^2(1-\rho^2) + (\sigma_{F,C} + \rho a)^2}{a^2(1-\rho^2) + (\sigma_{I,C} + \rho a)^2} < \frac{(\sigma_{F,C} + \rho a)^2}{(\sigma_{I,C} + \rho a)^2} \\ \Rightarrow\ r &< \frac{\sigma_{F,C} + \rho a}{\sigma_{I,C} + \rho a} < \frac{\sigma_{F,C}}{\sigma_{I,C}}, \quad \text{if } \rho > 0. \end{aligned} \tag{5}$$

Thus, we obtain the result that $\sigma_{F,C}/\sigma_{I,C}$ is bounded below by the measured $\sigma_{F,T}/\sigma_{I,T}$ when $\rho > 0$. Such values for ρ are highly plausible, and it is in fact difficult to imagine forces that would influence consumer error and analyst error in different directions; common unobservables (e.g., shocks) would lead to positive values for ρ. Still more general and precise results are possible: if we do not presume that $\sigma_{I,A} = \sigma_{F,A}$, differentiation and algebra show that the desired boundedness result holds when $\rho > -\frac{1}{2}[\sigma_{F,A}/\sigma_{F,C} + \sigma_{I,A}/\sigma_{I,C}]$, which again is always negative, but not necessarily less than -1. In terms of estimation using observed choices and covariates, we therefore know that *the measured value of* $\sigma_{F,T}/\sigma_{I,T}$ *provides a lower bound for the* (*unobservable*) *temporal stochastic inflation factor*, $\sigma_{F,C}/\sigma_{I,C}$, so long as the correlation in analyst and consumer errors is nonnegative (and in fact for any value of $\rho > -\frac{1}{2}[\sigma_{F,A}/\sigma_{F,C} + \sigma_{I,A}/\sigma_{I,C}]$). We thus concern ourselves throughout with measuring the total error ratio, $\sigma_{F,T}/\sigma_{I,T}$, and claim it as a lower bound for $\sigma_{F,C}/\sigma_{I,C}$.

We next present a statistical model for choices made among different items and times. Thus far, we have been as general as practicable about error distributions. They did not, for example, need to be independent, drawn from the same family, or even have the same mean. In our empirical applications, however, it is important to ensure that manipulations of error variance do not also distort other values in the model. We must take care to ensure that scaling of errors alters only variances, not means, thus allowing for cross-model comparison and interpretation.

Empirical Model

Consider an individual who is choosing a single item from a set of alternatives. The (conditional) multinomial logit (MNL) model has been used extensively in marketing and decision theory to model individual choice behavior (McFadden 1974), and MNL assumes the degree of uncertainty about individuals' anticipated utility is the same for all alternatives and time periods. This particular assumption limits our ability to model variations in preference uncertainty in observed choice patterns. For example, a consumer choosing from amongst a set of snack brands has varying degrees of *uncertainty* about how much she will enjoy each snack, irrespective of her best guess (i.e., mean anticipated utility). She feels relatively certain about how much she will enjoy the snacks she has the most experience eating and less so about the other snacks, especially snacks she has never seen before, irrespective of whether she believes she will like them. The analyst using MNL to model her choice behavior would be presuming equal degrees of predictability about each of these alternatives, ignoring that uncertainty varies in a meaningful, *estimable* way across them. Disregarding differences in brand-specific variation when they are present would not only be an a priori misspecification; it could induce artifacts into the stochastic variance effects whose verification and accurate measurement are our main interest. Imposing equal degrees of stochastic variation on each brand also entails the independence of irrelevant alternatives (IIA) property. Bhat's (1995) heteroscedastic extreme value (HEV) model and Allenby and Rossi's (1999) diagonal-covariance probit, both explored in the sequel, overcome this well-known problem without a huge penalty in parsimony (for further discussion, see also Baltas and Doyle 2001). Thus, although our focus throughout is on temporally induced variance effects, all models will accommodate brand-specific variation parameters as well, unless they are shown to be statistically nonsignificant.

One approach that allows for different degrees of stochastic variation involves the heteroscedastic extreme value model (Allenby and Ginter 1995, Bhat 1995). For each alternative, j, in the HEV model, the $\{\varepsilon_{ijt}\}$, for individual i at time t, follow a type 1 extreme value distribution with scaling parameter θ_j. The error terms, $\{\varepsilon_{ijt}\}$, are independent but not identically distributed, with variances equal to $\pi^2\theta_j^2/6$. The variance-scaling parameter, θ_j, is directly proportional to the standard deviation of the stochastic component of utility, ε_{ijt}. HEV nests MNL, with $\theta_j = 1$ for all j. The multinomial probit (MNP) model is yet more general, allowing multivariate normal (MVN) errors with arbitrary (estimable) error covariances.

To address the problem at hand requires that we incorporate an additional component into the model framework: uncertainty associated with time duration between choice and consumption. Specifically, we must allow individuals' uncertainty about anticipated utility to vary according to whether choices are made for immediate versus future consumption, so that the degree of stochastic variation can vary across choice conditions, as well as by brand/item. To this end, an additional variance-scaling parameter, σ_c (where c represents consumption delay, future or immediate), can be included in the random utility formulation,

$$U_{ijt} = V_{ijt} + \theta_j \sigma_c \xi_{ijt}, \tag{6}$$

where ξ_{ijt} is a univariate error draw. The specification (6) recognizes and allows estimation (using standard tools) of the variation in the stochastic component of utility associated with two distinct sources: that specific to the brand (θ_j), and that associated with the time between choice and consumption (σ_c). The specific multiplicative format we use, $\theta_j\sigma_c\xi_{ijt}$, is consistent with an additive-in-logs specification typical in modeling variance effects. Although other specifications are possible, we have not systematically explored them here (notably, a linear-additive formulation would require a number of estimation constraints to avoid negativity and would not naturally allow the baseline variance, brand-specific multipliers and temporal stochastic inflation parameter to interact as they do in (6)). Louviere and Eagle (2006) provide a full discussion of the role and specification of such scale constants.

Deterministic components, V_{ijt}, are expressed as linear-additive functions of brand-specific characteristics, X_{ijk}, that account for the utility of brand j for individual i. The full model is, therefore,

$$U_{ijt} = \sum_{k=1}^{K} \beta_k X_{ijk} + \theta_j \sigma_c \xi_{ijt}. \tag{7}$$

This formulation (7) enables us to rigorously test for many possible effects *over and above* others: the relative magnitude of uncertainty across available options (θ_j), uncertainty stemming from consumption delay (σ_c), and, via the error correlations possible with MNP, whether the stochastic portions of brand utilities can be treated as independent. We discuss the incorporation of consumer preference heterogeneity below and in our empirical application.

As discussed previously, we expect the temporal scaling factor, σ_c, to be greater when choosing for future time periods. With the convention that *FUT* refers to all choices made for consumption in future time periods, and *IMM* refers to choices made for consumption in the immediate time period, we posit

Hypothesis 1 (H1). $\sigma_{FUT}/\sigma_{IMM} > 1$.

As discussed by Louviere (2001), our analysis will consider the *same* deterministic specification across temporal conditions and ascertain whether equal variances or unequal ones explain empirical data patterns better. This analysis will lead to various estimates for the critical quantity, $\sigma_{FUT}/\sigma_{IMM}$, which as discussed earlier provides a lower bound for the consumer's stochastic inflation ratio.

We examine the effects of consumption time delay with a controlled laboratory experiment that varies the time between choice and consumption. In all discrete-choice models, various elements must be fixed to set location and scale (described fully in the sequel; for further discussion of relative error scaling, model specification, and identification for this and related classes of models, see Swait and Louviere 1993 and Louviere et al. 2000). Specifically, the scale of the error is confounded with the magnitude of the deterministic component. It is important to note that we alter the ratio of stochastic components *only* across time conditions (*FUT* versus *IMM*); model identification is set identically across time conditions otherwise.

We test for all effects by comparing nested models using standard likelihood-ratio tests, subject to three well-known identification constraints (see Bhat 1995): (1) one of the brand-specific constants (β_k) must be fixed, ordinarily to zero; (2) one of the brand-specific scaling parameters (θ_j) is set (arbitrarily) to one; and (3) one of the temporal scaling parameters is set to one (i.e., $\sigma_{IMM} = 1$). Therefore, the key test of H1 concerns whether $\sigma_{FUT} > 1$. We will also test for a number of other effects, as suggested by prior literature on discrete choice and time delay. The two key issues can be stated simply: is there evidence of temporal variance inflation (σ_c) *over and above* other effects, and is there evidence of other effects *over and above* temporal variance inflation?

Testing for Nonconstant Stochastic Variation

Experimental research on time delay is abundant, and it is not our goal to recap it here or to elucidate the drivers of all consequent behaviors. Rather, we wish to replicate aspects of a carefully conducted study in a manner that isolates possible temporal stochastic inflation effects, without identifying their underlying causes, and ruling out certain confounding factors. Perhaps the best known among such studies is Simonson's (1990), which found that people choosing multiple items from an assortment tend to choose a greater variety of items when all items are chosen "simultaneously" before (future) consumption versus when each item is chosen "sequentially" for immediate consumption. Similar differences in observed variety have been replicated by other researchers as well (e.g., Read et al. 2001, Read and Loewenstein 1995). Numerous explanations have been posited for the underlying psychological mechanism, although Loewenstein (2001) suggests that none yet offer a thorough account. Among the most widely cited explanations is Simonson's (1990) contention that the difference in variety chosen is due to greater uncertainty of future preferences for participants in the simultaneous condition. We build on this work using a methodology that allows for the possibility that uncertainty differences may arise because time delay can inflate the (unobserved) stochastic variation in utility. The forthcoming experiment offers a platform to detect and measure temporal inflation in (relative) stochastic variance.

Experiment

We conducted an experiment to assess the effect of consumption time delay on stochastic variation. The experimental design replicates essential features of Simonson (1990, Experiment 2) and of Read and Loewenstein (1995, Experiment 1), with one key difference: because preference heterogeneity is critical in accounting for choice, we collected measures of individual participants' preferences *before* they made any choices. This allowed us to directly test the effect of time delay on the probability of choosing *participant-specific* most- and least-favored alternatives, the quantities about which the random utility formulation makes sharp predictions.

Method

Design. A between-subjects design was used to test the effect of temporal distance on stochastic variation in utilities and choice probabilities. Participants chose three snacks from among a set of six. Snacks were selected to accord with those used by Simonson (1990) and Read and Loewenstein (1995). They included the following: Austin cheese crackers, Doritos tortilla chips, a Hershey's chocolate bar with almonds, Oreo cookies, Planters peanuts, and a Snickers bar. One hundred three undergraduate students participated to earn credit in an introductory business course. In addition, participants who completed all four parts of the study were given a $3 completion bonus.

Procedure. The experimental procedure comprised four consecutive sessions, each conducted one week apart. Four groups of 24–28 participants took part in the experiment. Each participant group was run as an independent set of 4 sessions, so that 16 sessions were conducted in total. Participants within a condition reported on the same day of week and time of day for each session of the experiment. In session one, participants' preferences for the available snack alternatives were measured by asking participants to rate how much they liked each snack, using an 11-point Likert scale ("1" = dislike very much, "11" = like very much). Choices were made during sessions two, three, and four (hereafter referred to as choice weeks one, two, and three, respectively).

In choice week one, participants read the following instructions describing the task to be completed. Only the "simultaneous" participants saw the words in parentheses, whereas only the "sequential" participants saw the words in square brackets.

> Today (the session a week from today, and the session two weeks from today), we will be giving away to students snacks of their choice. [Each student can choose any one of the items from the selection of snacks available on the table.] (Each choice can be any one of the items from the selection of snacks displayed on the table, and you may select the same snack more than once.) There is a sufficient supply of all snacks.

Participants in the *sequential condition* chose one snack to eat, wrote the name of their chosen snack on the same page as the instructions, and ate the snack immediately. All available snacks were displayed on a table, and participants walked to the table to pick up the snack they chose. Participants in the *simultaneous condition* chose one snack to eat immediately, a second snack to eat one week later (in choice week two), and a third snack to eat two weeks later (in choice week three). They wrote the name of each snack chosen on three separate sheets of paper, with each page designating a specific choice week in the experiment (i.e., choices were assigned to specific time periods). After making the three choices, participants ate their first chosen snack. Note that requiring simultaneous participants to preassign each choice to a specific consumption time period eliminated any opportunity to use variety as a "hedge" against changing preferences (Kreps 1979, Walsh 1995).

In choice week two, participants in the sequential condition performed the same task as in choice week one. All available snacks were again displayed and participants picked up their chosen snack from the table. Participants in the simultaneous condition were given the sheet of paper from choice week one, on which they had written the snack they chose to eat for that day. They selected the snack from the table and ate it. In choice week three, all participants performed the same task as in choice week two.

Choice Model Specification. To assess the effects of consumption time delay on participants' uncertainty about anticipated utility, we adopt a model that accords closely with those used in prior literature that did not allow for differences in stochastic variance across brands or time periods (Simonson 1990). Latent utility, U_{ijt}, for individual i, brand j, and time t, is specified as

$$U_{ijt} = \beta_{RATE}\, RATE_{ij} + \beta_j\, BRAND_j + \theta_j \sigma_c \xi_{ijt}, \quad (8)$$

where β_{RATE} represents the effect of prior rating on utility, $RATE_{ij}$ is individual i's prior rating for snack j, β_j represents the brand-specific constant for snack j, $\{BRAND_j\}$ are 0–1 brand dummy variables, and the three key stochastic constructs—brand scaling (θ_j), temporal scaling (σ_c), and error correlations—are as described previously. Given that simultaneous participants made their first choice for consumption in week one, and the remaining choices for future weeks, σ_{FUT} corresponds to all but the first choice in the simultaneous condition. The temporal scaling factor, σ_{FUT}, is therefore akin to an "average" inflation factor across time periods two and three. This could be easily generalized to multiple future inflation factors, albeit at the expense of parsimony. Thus, the model specification is

$$\begin{aligned} U_{ijt} = {} & \beta_{RATE} RATE_{ij} + \beta_{CHEESE} CHEESE \\ & + \beta_{DORITOS} DORITOS + \beta_{HERSHEY} HERSHEY \\ & + \beta_{OREO} OREO + \beta_{PEANUT} PEANUT + \theta_j \sigma_c \xi_{ijt}, \end{aligned} \quad (9)$$

where β_{CHEESE} and $CHEESE$ represent the brand-specific constant and 0–1 brand dummy variable, respectively, for Austin cheese crackers, and likewise for the remaining snacks. Note that although (9) comprises all available brands, brand dummies are included for all brands but one, for identification purposes; Snickers was arbitrarily selected to have its brand dummy set to zero. All models were estimated by iteratively combining routines in commercially available software packages (LIMDEP, MATLAB, and Stata) and checked by reprogramming the likelihood calculations using Gauss-Laguerre quadrature. Preliminary Bayesian analysis supported the presumed asymptotic normality of parameters' marginal densities, on which all our classical likelihood-ratio tests will be based, as well as accounting for observed heterogeneity (via *RATE*) in place of an additional random coefficients specification for unobserved heterogeneity. That is, the inclusion of *RATE* allows for so-called "observed heterogeneity"; "unobserved" (i.e., parametric) heterogeneity cannot be reliably accommodated because of the small number of (correlated) observations per subject, as cautioned by Andrews et al. (2008).[3]

[3] However, because Louviere and Eagle (2006), Louviere and Meyer (2007), and Adamowicz et al. (2008) caution against ignoring various sources of unobserved heterogeneity—including

Table 1 Empirical Model Comparisons

	LL	df	σ_{FUT}	*VS*	Likelihood-ratio tests: *p*-value versus base models			
					MNP	MNP-B	MNP-T	MNP-BT
Base models								
MNP: No brand or temporal scaling	−408.25	6	—	—	—	—	—	—
MNP-B: Brand, but no temporal scaling	−402.43	11	—	—	**0.040**	—	—	—
MNP-T: Temporal, but no brand scaling	−397.46	7	2.11	—	**0.000**	—	—	—
MNP-BT: Brand and temporal scaling	−390.57	12	2.37	—	0.000	**0.000**	0.017	—
MNP-BF: Brand and free temporal scaling	−385.93	17	2.53[a]	—	0.000	0.000	0.011	**0.100**
With correlated error terms								
MNP + EquiCorrelation	−408.20	7	—	—				
MNP-B + EquiCorrelation	−401.55	12	—	—				
MNP-BT + EquiCorrelation	−390.25	13	2.44	—	0.000	0.000	0.025	**0.425**
MNP + Free correlation	−402.10	16	—	—				
MNP-B + Free correlation	−400.29	21	—	—				
MNP-BT + Free correlation	−388.56	22	2.70	—	0.001	0.004	0.274	**0.947**
Allowing different preference weightings								
MNP + *RATE* ∗ *SIM* interaction	−406.39	7	—	—	0.053	—	—	—
MNP-B + *RATE* ∗ *SIM* interaction	−399.96	12	—	—	0.011	0.027	—	—
MNP-T + *RATE* ∗ *SIM* interaction	−396.84	8	2.33	—	0.000	—	0.268	—
MNP-BT + *RATE* ∗ *SIM* interaction	−390.07	13	2.56	—	0.000	0.000	0.022	**0.318**
Variety-seeking/inertia models								
MNP + Variety-seeking	−407.25	7	—	−0.043	0.158	—	—	—
MNP-B + Variety-seeking	−402.18	12	—	−0.032	0.059	0.487	—	—
SIM only[b]	−202.19	12		−0.001				
SEQ only	−177.70	12		−0.125				
χ^2, $\Delta(VS) = 2.184$, $p = 0.139$								
MNP-T + Variety-seeking	−397.01	8	2.10	−0.031	0.000	—	0.344	—
MNP-BT + Variety-seeking	−390.38	13	2.38	−0.021	0.000	0.000	0.028	**0.538**

Notes. Bold denotes key statistical tests referred to in the text.
[a]The harmonic mean of $\sigma_{FUT,j}$ is listed for "free temporal scaling."
[b]Significance test for *SIM VS*: $p = 0.995$; for *SEQ VS*: $p = 0.060$.

The utility-based model (8) allows us to evaluate *marginal* explanatory power: whether a particular effect is supported over and above others. The large number of distinct deterministic and stochastic effects our formulation affords can lead to a combinatorial explosion in possible models. Table 1 summarizes the range of relevant comparisons and tells a clear story of patterns in the choice data. First, however, a practical matter concerns whether the more tractable logit-based (MNL or HEV) model or the probit (MNP) model is more appropriate. We estimated both logit-based and probit-based models and found identical patterns of results. Because in every case the probit specification offered better fit, we formally present the probit model results exclusively but report informally that effect strengths were nearly identical under the logit specification (except those with error correlations, which the standard logit does not support).

unobserved coefficient and scale heterogeneity—we reestimated many of the key models in this article subject to a variety of fixed, reduced coefficient values for b_{RATE}; these resulted in mild decreases in the estimated value of σ_c. Data with substantially more choices per individual would be required to settle this issue empirically.

Results

Before presenting our focal results, we first confirm that those of prior research are replicated. As expected, simultaneous participants chose a significantly greater number of unique items than participants choosing sequentially ($M_{SIM} = 2.42$, $M_{SEQ} = 1.82$, $p < 0.001$). This result confirms previous research (Read and Loewenstein 1995, Simonson 1990) showing that simultaneous choice leads to a relatively greater number of observed unique items than sequential choice. Next, we measure the degree of relative temporal stochastic variance effects, and in so doing assess H1.

For simplicity of nomenclature in presenting all forthcoming results, the term "uncertainty" refers to the degree of stochastic variation.

Temporal Differences in Uncertainty. Hypothesis 1 suggested that stochastic variation in choices made for future consumption would be significantly greater than those made for immediate consumption ($\sigma_{FUT} > 1$). Using the convention that "B" denotes brand scaling and "T" temporal scaling, comparing model "MNP-BT" to (the nested) model MNP-B (i.e., imposing the restriction that $\sigma_{FUT} = \sigma_{IMM} = 1$)

indicated a strongly significant difference ($\chi^2_{\text{diff}}(1) = 11.86$, $p < 0.001$; see Table 1).[4] Even when allowing for brand-specific stochastic variation, the temporal scaling parameter estimate was $\sigma_{FUT} = 2.37$, so that relative stochastic variation is approximately 137% greater when choosing for future (versus immediate) consumption. Recall that the ratio of total error variances is a lower bound for the ratio of consumer error variances, and thus 2.37 is a conservative estimate of the consumers' temporal inflation factor. We report parenthetically that this finding was fairly robust to error specification, as evidenced by testing H1 assuming a logit model ($\chi^2_{\text{diff}}(1) = 9.60$, $p < 0.001$; $\sigma_{FUT} = 2.33$).

This evidence supports H1 taken on its own. Moreover, and even more compelling, allowing different degrees of stochastic variation (σ_c) for future and for immediate consumption was supported in *every* context, no matter which other effects were present, and this is among the main findings of our analysis. A key issue concerns the *degree* of temporally induced inflation. For all models that include it, the estimated value of σ_{FUT} is remarkably consistent. As per Table 1, it falls between 2.10 and 2.70 and, even when included in models already allowing for brand-specific scaling, is significant at $p < 0.001$, suggesting that temporal inflation is a strong, robust effect *over and above* any others included in the model.

Brand-Specific Differences in Uncertainty. Although not a main focus of our investigation, it is reasonable to ask whether allowing stochastic variation to differ across brands just adds needless complexity. We find that imposing the restriction that θ_j takes a common value (across brands, j) leads to a worse fit. Likelihood-ratio tests comparing model MNP-BT to (the nested) model MNP-T ($\chi^2_{\text{diff}}(5) = 6.89$, $p < 0.02$) support that variation does differ across brands, over and above any between-condition temporal scaling. Furthermore, allowing brands to have differing degrees of stochastic variation was *always* supported, over and above any other effects in the model (ps < 0.05; brand-scaling parameter estimates, θ_j, along with all supplemental estimation results, are available from the authors). We therefore include brand-specific stochastic variation (uncertainty) in all model comparisons discussed below.

Brand-Specific Differences in Temporal Scaling. The most general hypothesis involving temporal inflation is that each brand's variance is inflated in future choice, but not to the *same degree*. That is, each brand j has its *own* temporal inflation factor, σ_{cj}, rather than a common temporal inflation factor, σ_c. This is tested in Table 1 and designated as "free temporal scaling" (MNP-BF). Although strongly supported on its own ($\chi^2_{\text{diff}}(11) = 22.32$, $p < 0.001$), it is *not* supported over and above there being a *single* value for temporal inflation ($\chi^2_{\text{diff}}(5) = 4.64$, $p \approx 0.10$). In other words, time delay between choice and consumption revealed a consistent, inflationary effect on uncertainty about future experienced utility for each of the choice alternatives. We note parenthetically, however, that the estimated values of σ_{cj} varied in a pattern resembling an inverted U-shape relative to mean brand rating: σ_{cj} values of $\{1.76, 2.74, 4.93, 4.29, 2.81, 1.00\}$ corresponded to mean ratings $\{5.53, 6.78, 6.83, 8.09, 7.99, 8.55\}$ and choice shares $\{4.2\%, 6.2\%, 10.7\%, 17.9\%, 18.2\%, 42.7\%\}$, respectively. This suggests the intriguing possibility that brands that are strongly preferred or weakly preferred may be subject to relatively less temporal inflation. We must bear in mind, however, that statistical support for this contention would likely require far larger samples, given the weak degree of evidence in its favor here ($p \approx 0.10$).

Error Correlations. It is well known that although models like MNL are convenient tools, those allowing for correlated errors are far more demanding and, for want of a better term, finicky in applications (Harding and Hausman 2006). Nevertheless, it is important to explore whether the pattern of experimental results could be driven by latent error correlations, and we consider two possible patterns: a highly parsimonious one where all errors are intercorrelated to the same degree ("equicorrelation"), and "free correlation," where error correlations are completely unrestricted. As shown in Table 1, neither equicorrelation ($\chi^2_{\text{diff}}(1) = 0.32$, $p > 0.4$) nor free correlation ($\chi^2_{\text{diff}}(10) = 2.01$, $p > 0.9$) was supported. This suggests that the proposed model, accounting for brand and temporal scaling (but no other error effects), is adequate to capture stochastic variation patterns in the choice data.

Differences in the Effect of Prior Preference on Choice. Previous literature has posited that underlying preferences are underweighted in simultaneous choice. To assess this, we allow the coefficient on *RATE* to differ by condition, thereby allowing an interaction between prior preference and choice condition, $RATE_{ij} * SIM$, where *SIM* is a binary dummy that equals one for the simultaneous condition and zero for the sequential condition. Intriguingly, consistent with previous research, this interaction was significant in models that did *not* allow temporal scaling: for the standard probit model (MNP; $\chi^2_{\text{diff}}(1) = 1.84$, $p = 0.053$) and the probit with brand scaling (MNP-B; $\chi^2_{\text{diff}}(1) = 2.47$, $p < 0.03$). However, it was *never* significant for

[4] Given the numerous model comparisons reported using chi-square, we adopt a notation convention of listing the difference in log-likelihood between models (rather than two times that quantity); i.e., $\chi^2_{\text{diff}}(\Delta df) = \Delta\text{LL}$. This serves to facilitate comparing results reported within the text to those appearing in Table 1.

models including temporal scaling (MNP-T, $\chi^2_{\text{diff}}(1) = 0.62$, $p > 0.2$; MNP-BT, $\chi^2_{\text{diff}}(1) = 0.50$, $p > 0.3$). The reverse is not true: temporal scaling was strongly significant when added to *any* model including the $RATE_{ij} * SIM$ interaction term (MNP versus MNP-T, $\chi^2_{\text{diff}}(1) = 9.55$, $p < 0.001$; MNP-B versus MNP-BT, $\chi^2_{\text{diff}}(1) = 9.89$, $p < 0.001$). In other words, once temporal variation in uncertainty is taken into account, there is no evidence that simultaneous and sequential participants weight underlying preferences differently during choice. Critically, one might erroneously conclude the opposite if temporal differences in stochastic variation are not accounted for in the analysis.

Finally, we tested whether the deterministic component of utility is itself stable over time by estimating a model that nests MNP-BF. Six additional parameters were introduced to allow b_{RATE}, as well as brand-specific constants, to differ for *FUT* and *IMM*, but did not significantly improve fit over MNP-BF ($p \approx 0.08$).

State Dependence/Variety Seeking. Patterns of repeated choice can display various "nonzero order" behavior in the form of state dependence, habit persistence, or variety seeking; these can be confounded with error-scaling effects. Seetharaman (2004) proposed a utility-theoretic account of four different sources of state dependence—lagged choices (structural state dependence), serially correlated error terms or choices (habit persistence types I and II), and lagged covariates—finding the first to be the most important. As advocated in that paper, we apply a specific utility-based model (Seetharaman and Chintagunta 1998) to recover variety-seeking effects and reestimated all models subject to their hybrid variety-seeking/inertia specification:

$$P_{ij} = \frac{|VS| + VS}{2}\left(\frac{P_j}{1 - P_i}\right) + (1 - |VS|)P_j \quad (\text{for } i \neq j);$$
$$P_{jj} = \frac{|VS| - VS}{2} + (1 - |VS|)P_j, \tag{10}$$

where P_{ij} is the conditional probability of choosing alternative j for the next time period (given that i was chosen for the previous period), and P_j is the unconditional choice probability resulting from Equation (8). The parameter VS, ranging from -1 to 1, represents the degree of inertia or variety seeking in the data, with negative values indicating inertia.

As shown in Table 1, VS was negative, although nonsignificant, for *all* models, suggesting a slight degree of inertial choice behavior (for example: MNP-BT; $VS = -0.021$, $\chi^2_{\text{diff}}(1) = 0.19$, $p > 0.5$). To assess this as a driver for variety differences, however, we reestimated the model separately for the simultaneous ($VS_{SIM} = -0.001$, $p > 0.9$) and sequential ($VS_{SEQ} = -0.125$, $p > 0.06$) conditions. As expected, there appeared to be *less inertia* in the simultaneous condition, but the difference was not significant ($VS_{\text{diff}} = 0.124$, $\chi^2 = 2.184$, $p > 0.13$); it would seem inappropriate to interpret this as "more variety seeking," because choices exhibited no variety seeking in either condition. Finally, temporal scaling was strongly significant when added to any model allowing for variety seeking (MNP versus MNP-T; $\chi^2_{\text{diff}}(1) = 10.24$, $p < 0.001$; MNP-B versus MNP-BT; $\chi^2_{\text{diff}}(1) = 11.80$, $p < 0.001$). Taken together, these results indicate that, although average observed variety was greater in the simultaneous condition, there is no evidence that simultaneous decision makers engage in more variety seeking.

Summary of Key Findings. In summary, our pattern of results underscores the importance of examining, and testing for, both deterministic and stochastic effects on choice behavior. Whenever our stochastic variation constructs, brand and temporal scaling, are added to models including the other posited effects, they are *always* strongly supported (ps < 0.001). However, when other potential effects—differential preference weighting, correlated errors, or variety seeking—are added to the proposed model (with brand and temporal scaling; MNP-BT), *none* of these effects is supported (ps > 0.10; see the far right column of Table 1). The data therefore tell a surprisingly direct and parsimonious story: different brands entail different degrees of stochastic variation, there is greater stochastic variation for choices made for future than for immediate consumption, and, moreover, this can be adequately captured by a single temporal variance inflation factor.

We have thus far appealed to real data to demonstrate the value of accommodating both deterministic and stochastic effects when modeling choice. However, real data, even in controlled experiments, can contain confounds that interfere with causal inference. Critically, we can never fully determine the effects of model misspecification: can a model lacking some particular feature lead to an incorrect interpretation of one's data? To examine this issue, we next turn to simulated choice scenarios whose generating mechanisms—and whether they contain either deterministic or stochastic effects—are known.

Simulation Study 1: Presuming Constant Stochastic Variance as Potential Specification Error

Having measured the extent of temporal stochastic inflation, we can now investigate a question of key importance: could common analytical tools *erroneously* estimate effects in such a way as to lead to an unsubstantiated conclusion? A Monte Carlo simulation study allows us to examine the extent to which deterministic effects can be either correctly statistically detected or incorrectly inferred due to unfounded assumptions about the stochastic portion of utility. We therefore examined two hypothetical choice scenarios, based in part on our experimental design and empirical findings—one in which simultaneous versus sequential choice differences are driven by stochastic differences only, and a second in which choice differences are instead rooted in deterministic differences. In our analysis of each scenario, we examine three factors that influence choice (akin to a $2 \times 2 \times 2$ experimental design): brand-specific uncertainty (fixed θ_j versus unrestricted θ_j), temporal differences in uncertainty (fixed $\sigma = 1$ versus $\sigma \geq 1$), and differences in preference weighting across choice conditions ($b_{RATE*SIM} = 0$ versus $b_{RATE*SIM} < 0$). Previous research posited that sequential decision makers are more likely to choose their favorite option, and less variety, because they weight preferences more heavily ($b_{RATE*SIM} < 0$) than simultaneous decision makers (Simonson 1990). This "deterministic hypothesis" can be tested with both prior preference (rating) and a preference interaction term (with coefficient $b_{RATE*SIM}$) as covariates in Equation (8) or Equation (9). Thus, our scenario analyses will directly measure and test for both stochastic and deterministic influences on choice in each scenario.

The two choice scenarios simulate simultaneous versus sequential choice processes as follows:

(1) *Stochastic Scenario*: Temporal ($\sigma > 1$ with time delay) and brand ($\{\theta_j\} \neq 1$) differences in uncertainty, no difference in preference weighting ($b_{RATE*SIM} = 0$).

(2) *Deterministic Scenario*: Neither temporal ($\sigma = 1$) nor brand ($\{\theta_j\} = 1$) differences in uncertainty, difference in preference weighting ($b_{RATE*SIM} < 0$).

These scenarios were carefully chosen to examine the key issue in our study: the effects of temporal scaling in choice. Both are described and generated by parameters of the discrete-choice model, as follows. In stochastic Scenario 1, prior preferences are *not* weighted more strongly in the sequential condition. That is, Scenario 1 allows for stochastic variance effects, but a classic (deterministic) explanation of variety is *not* present. In deterministic Scenario 2, there are *no* stochastic variance effects, but there *is* an interaction effect representing a systematic shift in preference weighting across choice conditions. For both scenarios, we will estimate a range of models to quantify how various empirical features of the data could lead to statistical artifacts. Specifically, we explore how a researcher might reach *incorrect* conclusions: claiming there is an underlying (deterministic) shift in preferences when there is not (possible in Scenario 1), or declaring a systematic difference in stochastic variation when there is not (Scenario 2).

Data Generation. Our choice scenarios simulate experimental subjects choosing three snacks from among a set of six, similar to our experimental design described earlier. The simulation study relies on generated data, as specified fully below. Conceptually, we must distinguish the pattern in the underlying *covariate* data (subject-specific preferences for the various brands) and in the resulting *choice* data. Although demonstrating our main theoretical contentions via arbitrarily chosen simulation parameters would not have been difficult, we wish both the covariate and choice data to reflect empirically meaningful patterns. Thus, for the simulated covariates (preferences), we use the preference data from our experiment; for the choice data, we generate them via known statistical processes and parameter values stemming from our analysis of choices made in the experiment.

The simulated choice and preference data were generated in a manner consistent with a multivariate ordered probit model (Lawrence et al. 2008). First, we estimated the latent (MVN) distribution underlying the collected brand preferences, using polychoric correlations and maximum likelihood estimation (see Jöreskog 1994). Estimated cutoffs converted draws from this MVN distribution to ordinal data, so that the simulated and real probability density functions (pdfs) matched (e.g., the proportion rating Oreo a "7" out of a possible 11 was identical for the simulated draws and real preference data). In this way, we can capture respondent scale usage and the correct (marginal) preference distributions for each brand (as well as their intercorrelations). A final advantage to this approach is that we can completely control the process generating the *outcome* variable: brand choice, both over time and across choice conditions. Specifically, using carefully selected discrete-choice model parameters, we can generate outcomes perfectly consistent with a hypothesized model of choice behavior, and examine the "empirical" effects of various assumptions, temporal and otherwise. We can thereby determine what is truly statistically artifactual, an impossibility using field data on actual choices.

We control how choices themselves are generated, as follows. Given our coefficients, and covariates drawn from the master generated list, we calculate a (latent) deterministic utility for each of the six brands on any particular choice occasion. To these, we add a random (normal) error draw, scaled by condition and choice occasion according to (8). "Choice" on a particular occasion is then simply the brand with the largest realized utility. This is carried out for both the simultaneous (*SIM*) and sequential (*SEQ*) "conditions" using the *same* covariate draws (i.e., simulated respondents) to eliminate this as a source of between-condition differences. Error draws are made separately across conditions, as they must be. Our

Table 2 Simulation Study 1: Probit-Based Model Comparisons for "Stochastic" and "Deterministic" Scenarios

	Estimated parameters					Tests against null of		
Model	σ	$\{\theta_j\}$	$b_{RATE*SIM}$	No. of parameters	LL	$\sigma=1$	$\{\theta_j\}=1$	$b_{RATE*SIM}=0$
"Stochastic scenario" Temporal scaling ($\sigma>1$), Brand scaling ($\{\theta_j\}\neq 1$), No preference interaction ($b_{RATE*SIM}=0$)								
M1	1	1	No	5	−4,059.6	—	—	—
M2	1	1	Yes	6	−4,014.7	—	—	E-21
M3	1	Free	No	10	−4,000.0	—	E-24	—
M4	1	Free	Yes	11	−3,948.2	—	E-27	E-24
M5	2.08	1	No	6	−3,961.6	E-44	—	—
M6	2.00	1	Yes	7	−3,960.8	E-25	—	0.197
M7	2.38	Free	No	11	−3,862.9	E-61	E-41	—
M8	2.38	Free	Yes	12	−3,862.9	E-39	E-40	0.879
"Deterministic scenario" No temporal scaling ($\sigma=1$), No brand scaling ($\{\theta_j\}=1$), Preference interaction ($b_{RATE*SIM}\neq 0$)								
M1	1	1	No	5	−4,051.9	—	—	—
M2	1	1	Yes	6	−4,021.5	—	—	E-15
M3	1	Free	No	10	−4,050.3	—	0.665	—
M4	1	Free	Yes	11	−4,020.2	—	0.754	E-15
M5	1.25	1	No	6	−4,042.6	E-05	—	—
M6	1.01	1	Yes	7	−4,021.5	0.840	—	E-11
M7	1.27	Free	No	11	−4,040.4	E-06	0.490	—
M8	1.01	Free	Yes	12	−4,020.2	0.892	0.757	E-10

simulated data sets consist of 500 draws from the master covariate list, yielding 500 sets of three choices for *each* of the *SIM* and *SEQ* conditions. Thus, our task will be to generate and statistically account for a total of 3,000 choices from among six snacks.

Discrete-choice models require identification constraints for brand-specific constants (we will use $b_{SNICKERS}=0$) and also for variances; to ensure that results are comparable across model specifications, we use the rescaling rule that the *product* of all brand-specific variances be one: $\Pi_j(\theta_j)=1$. Parametric values for the stochastic scenario are taken from model MNP-BT in Table 1, with $b_{RATE}=0.331$, $b_{RATE*SIM}=0$, $b_j=\{-0.869, -1.138, -0.370, -0.903, -0.695, 0\}$, $\theta_j=\{0.782, 1.336, 0.280, 1.070, 0.467, 1\}$, and $\sigma_{FUT}=2.37$; deterministic scenario parameter values are taken from "MNP + $RATE*SIM$ interaction" in Table 1, with $b_{RATE}=0.353$, $b_{RATE*SIM}=-0.112$, $b_j=\{-0.857, -0.576, -0.724, -0.565, -1.016, 0\}$, $\theta_j=\{1\}$, and $\sigma_{FUT}=1$. For our analyses of both simulated scenarios, we estimate the same set of eight parametrically nested models (M1–M8), in accord with the $2\times 2\times 2$ "analysis design" described previously. The eight models allow for varying degrees of flexibility. Models M1, M2, M5, and M6 do not allow for *brand* scaling, whereas the other four models do; nested models M5–M8 allow for *temporal* scaling to distinguish the *SIM* and *SEQ* conditions. Models M7 and M8 are the most general, and models M1 and M2 are the least.

Estimation and Empirical Results. Confronting the two simulated data sets with a variety of model specifications helps gauge the extent to which deterministic effects can be either correctly statistically detected or incorrectly inferred due solely to model assumptions that fail to match the true data generation process. Note that model M8 accounts for both types of scaling and the interaction, so we can, in theory, detect all three effects. All estimated quantities and resulting tests appear in Table 2, and we will refer to them throughout our discussion. For conciseness, we have isolated the most relevant tests here; full estimation results for these and other simulated data sets are available from the authors. We examine three classes of tests, as specified by their null hypotheses: temporal scaling ($\sigma=1$), brand scaling ($\{\theta_j\}=1$), and preference interaction effect (i.e., whether prior preferences are more heavily weighted in *SEQ* than *SIM*, with null $b_{RATE*SIM}=0$). Hypotheses are assessed via likelihood-ratio tests, rather than Wald-type t-tests, when possible.

Stochastic Scenario 1 simulated a choice process with differences in stochastic variation (uncertainty) for *SEQ* versus *SIM* choice, but no differences in preference weighting. As shown in Table 2, all effects—for brand scaling, temporal scaling, and interaction—are strongly significant ($p<$ E-21) in all models, with the exception of the two nested models (M6 and M8), where the preference interaction does not approach significance ($p>0.1$). Here is the main point: *A model that does not account for temporal scaling* (*such as the commonly applied M2*) *may lead to a conclusion that there is a difference in preference weighting, when there is not*. These effects are stark and unequivocal and speak directly to whether the standard framework for testing such hypotheses, in which error variance is identical across time, is appropriate. Note that testing certain behavioral theories, such as the diversification effect originally studied in this context, have relied on precisely this sort of test and underlying, constant variance, random utility model.

Finally, one might question whether *any* sort of tinkering whatever with the latent utility variance in the simulated data can lead to such incorrect inferences. This is *not* the case: when brand scalings do differ (in the "Stochastic" scenario) but are restricted to be identical in the analysis, this does not lead to such incorrect inferences. As per Table 2, M6 imposes (wrongly) that $\{\theta_j\} = 1$, yet the preference interaction is still correctly found to be nonsignificant ($b_{RATE*SIM} = -0.029$; $p > 0.1$).

The "stochastic" Scenario 1 simulation amply demonstrates the danger of ignoring temporal (though not necessarily brand) scaling when there is no true interaction. What happens, then, when there *is* an interaction but no scaling effects? Would the suggested model (M8) *mistake* the true interaction effect for some sort of variance scaling? Our second simulation, the "deterministic scenario," was set up to test this possibility.[5] Results appear in Table 2 and indicate that the unrestricted model (M8) is capable of explaining the data pattern well: both temporal and brand scaling are nonsignificant, whereas the interaction is strongly so ($p \approx$ E-10). Note that *all* models allowing for an interaction (M2, M4, M6, and M8) find extremely strong support for it in the data. Moreover, *all* models that allow for both an interaction and temporal scaling find the stochastic effects not to approach significance (ps > 0.5). Tellingly, M5 and M7, which incorrectly restrict the interaction coefficient to zero, do find small but significant estimated values of σ. Given the very large sample sizes and extreme significance levels for the interaction term, there is little question that interactions are *not* mistakenly "picked up" as stochastic effects when the remainder of the model is suitably general. Recall that our stochastic scenario simulation found exactly the opposite: that stochastic effects, *when unmodeled*, can appear to be strong interactions.

One might question whether we needed such large samples to validate temporal inflation. This can be answered readily using the "stochastic scenario" and either bootstrapping (i.e., resampling) or simply scaling our results and recalculating likelihood-ratio tests. Both procedures yield the same answer: all reported significant results for 500 simulated respondents remain so ($p < 0.001$) with only 50; the (nonsignificant) interactions yield p-values of about 0.8 with so few simulated respondents. Thus, the pattern of results reported here hold even with what would traditionally be termed small samples.

[5] Several "stochastic *and* deterministic scenario" simulations were run as well, but we do not discuss them here. They do, however, support the accuracy of the estimation methods, which recovered all parameters successfully.

Simulation Study 2: Verifying Robustness

We sought to test the robustness of our simulation findings with a supplementary simulation study based on the empirical results of Simonson (1990, Study 2). Would we find similar results assuming a different error distribution or with simulated choices based on different empirical analysis results? Reexamining Simonson's (1990) empirical findings offers such an opportunity. We examined the same two choice scenarios as in study 1—stochastic and deterministic—but instead generated simulated choices using the empirical model coefficients reported by Simonson in his seminal study (1990, Study 2). Thus, we have in both scenarios that $b_j = \{0.15, -0.10, 0.53, 0, -0.18, 0.45\}$. In the stochastic scenario: $b_{RATE} = 0.80$[6] and $b_{RATE*SEQ} = 0$, with $\sigma_{FUT} = 2.50$ to align with our empirical findings above. To avoid having to introduce new, arbitrary values into the simulation, brand-specific variances are scaled by this same inflation parameter; three were scaled up, two were unaltered, and one was scaled down. So that comparisons would be meaningful *across* models, we maintained the rescaling convention that $\Pi_j(\theta_j) = 1$, yielding $\theta_j = \{1.842, 1.842, 0.737, 0.737, 1.842, 0.295\}$; that is, as explained previously, $1.842/0.737 = 0.737/0.295 = 2.50$. For the deterministic scenario: $b_{RATE} = 0.60$, $b_{RATE*SEQ} = 0.40$, $\theta_j = \{1\}$, and $\sigma_{FUT} = 1$. That is, in the deterministic scenario data set, b_{RATE} and $b_{RATE*SEQ}$ are set such that preference is weighted by 0.6 in simultaneous choice, and increases to 1.0 for sequential choice, yielding the same average preference weighting (0.8) as in the stochastic scenario. Finally, note that in Simonson's data and for both simulations, the mean (i.e., brand

[6] Because Simonson (1990) used two presumably correlated preference measures, attractiveness (*ATR*) and liking (*LIK*), reported values must be combined to create a single usable coefficient, and the same must be done for *ATRSEQ* and *LIKSEQ*, which reflected how much greater these were in the sequential condition. These reported values were $b_{ATR} = 0.36$ in *SIM* and $0.36 + 0.39 = 0.75$ in *SEQ*; $b_{LIK} = 0.47$ and $0.47 + 0.21 = 0.68$, respectively. If these were to be the same across the *SIM* and *SEQ* conditions, the new means should be about $0.36 + (1/2)(0.39) = 0.555$ for b_{ATR} and $0.47 + (1/2)(0.21) = 0.575$ for b_{LIK}. In addition, because these will be correlated, they should not be used as separate covariates, and so we use $b_{RATE} = 0.8$ in the stochastic scenario (this corresponds to *ATR-LIK* correlation of approximately 0.55. Substantive model results were insensitive to various (positive) values of this correlation).

Table 3 Simulation Study 2: Model Comparisons for Scenarios Based on Simonson (1990) Estimates

Model	Model type	Estimated parameters: σ	$\{\theta_j\}$	$b_{RATE*SEQ}$	No. of parameters	LL	Tests against null of: $\sigma = 1$	$\{\theta_j\} = 1$	$b_{RATE*SEQ} = 0$
				"Stochastic scenario" Temporal scaling ($\sigma > 1$), Brand scaling ($\{\theta_j\} \neq 1$), No preference interaction ($b_{RATE*SEQ} = 0$)					
M1	Logit	1	1	No	5	−4,744.6	—	—	—
M2	Logit	1	1	Yes	6	−4,690.5	—	—	E-24
M3	HEV	1	Free	No	10	−4,642.7	—	E-41	—
M4	HEV	1	Free	Yes	11	−4,566.4	—	E-50	E-34
M5	Nested logit	2.63	1	No	6	−4,655.2	E-40	—	—
M6	Nested logit	2.50	1	Yes	7	−4,654.8	E-16	—	0.401
M7	Nested HEV	2.33	Free	No	11	−4,509.4	E-59	E-60	—
M8	Nested HEV	2.24	Free	Yes	12	−4,509.1	E-26	E-60	0.458
				"Deterministic scenario" No temporal scaling ($\sigma = 1$), No brand scaling ($\{\theta_j\} = 1$), Preference interaction ($b_{RATE*SEQ} \neq 0$)					
M1	Logit	1	1	No	5	−3,542.4	—	—	—
M2	Logit	1	1	Yes	6	−3,461.2	—	—	E-37
M3	HEV	1	Free	No	10	−3,537.9	—	0.108	—
M4	HEV	1	Free	Yes	11	−3,456.1	—	0.068	E-37
M5	Nested logit	1.35	1	No	6	−3,512.3	E-15	—	—
M6	Nested logit	0.97	1	Yes	7	−3,461.1	0.591	—	E-24
M7	Nested HEV	1.35	Free	No	11	−3,506.9	E-15	0.056	—
M8	Nested HEV	0.97	Free	Yes	12	−3,456.0	0.636	0.070	E-24

constant) and (stochastic) variance for Oreo is set to zero and to one for identification.

For our analyses of both simulated scenarios, we estimate the same set of eight parametrically nested models (M1–M8), similar to the analysis for the simulations just described above, but presume a Gumbel error distribution (as did Simonson). The eight models can be conceptualized in terms of shorthand names: Logit, HEV, Nested logit, and Nested HEV. The logit-based models do not allow for *brand* scaling, whereas the HEV models do; nested models allow for *temporal* scaling to distinguish the two "branches," the *SIM* and *SEQ* conditions. Thus, the "Nested HEV" model is the most general, the "Logit" the least. As shown in Table 3, the pattern of results replicates those found in the earlier simulation results. All stochastic scenario effects—for brand scaling, temporal scaling, and interaction—are strongly significant ($p <$ E-16) in all models, with the exception of the two nested models (M6 and M8), where the preference interaction does not approach significance ($p > 0.4$). The models that allow for an interaction but do not allow for temporal scaling effects—M2 and M4—erroneously indicate a strongly significant interaction effect ($p <$ E-24). Examining the deterministic scenario, the most general model (M8) explains the data pattern well: both temporal and brand scaling are nonsignificant, whereas the interaction is strongly so ($p \approx$ E-24). Again, we find that interactions are not mistakenly detected as stochastic effects when the model is suitably general.

In summary, our simulation results indicate that a model that does not account for temporal scaling may lead to an erroneous conclusion that there is a difference in preference weighting, when there is not. Conversely, a model that accounts for both temporal scaling and interaction effects ably disentangles deterministic and stochastic effects, avoiding that pitfall of misspecification. This result is robust across different presumed error distributions and different empirically based simulated choices.

Discussion and Conclusions

Ample evidence suggests that stochastic variation can differ across choice conditions. Marketing theory and research methods, however, in an effort to illuminate individual decision processes, often focus primarily on deterministic effects. Although such deterministic influences are doubtlessly important to explicate and verify, potentially stochastically driven effects are rarely accounted for in theory testing, as highlighted by Louviere (2001), who detailed specific contexts ripe for dedicated investigations. As such, it is presumed, usually tacitly, that any error "noise" and its specification cannot drive substantive implications for the phenomena under study. In this paper, we examined the validity of this assumption, using a well-established behavioral phenomenon posited to involve greater uncertainty for the future. Specifically, substantive conclusions can indeed arise artifactually if researchers fail to provide an account of both deterministic *and* stochastic components of utility.

Using a discrete-choice framework and carefully modeling the stochastic term to account for temporal differences in variation, our experiment and series of simulation studies provided strong empirical evidence that stochastic variation in anticipated utility can differ across temporal conditions. This evidence remained equally compelling when other purported (deterministic) explanations were included and made these other explanations recede to nonsignificance: aside from purely econometric issues like various patterns of error intercorrelation, substantive explanations involving state dependence, variety seeking, or greater weighting of underlying preferences in sequential choice were not supported.

Time delay offers an excellent example of a well-established phenomenon that has been posited to entail less "certainty" for future choices. However, it must be stressed that the present work makes no claims about the "simultaneous" decision-making process per se, and especially not that it is driven wholly by temporal stochastic inflation. Rather, our goal is to show that presuming such inflation away can lead to nontrivial substantive artifacts. A full account of the underlying *source* of inflated uncertainty would require ruling out several competing process explanations, over and above conceptualizing future uncertainty as comprising any and all contributory sources. Although we did not seek to identify or isolate any such alternative process explanations, our experiment did take care to rule many out, chief among them flexibility seeking (Kreps 1979, Walsh 1995), as well as uncertainty in time, location, and number of future consumption occasions. We were able, however, to quantify the overall effect, via a proof on the lower bound for anticipated future utility. To our knowledge, this thereby provides the first formal estimation of the magnitude (and significance) of the effect of time delay between choice and consumption. The degree of consistency of variance inflation measurements—between 2.10 and 2.70—across models is reassuring but should be supplemented by studies of other product classes, time delays, and choice set sizes. Such variance inflation factors are, of course, distinct in magnitude and concept from time-discounting parameters common in studies of intertemporal choice. Moreover, brand- and time-specific error-scaling parameters provide only two dimensions across which the degree of "unobserved variability" has been posited to contain structure. As detailed by Adamowicz et al. (2008), different contexts can evoke distinct choice processes and strategies, which may entail differing degrees of unobserved variability. In our view, this presents a ripe area for further investigation, particularly so when lab (e.g., conjoint) measurements are compared with those arising from market data.

Beyond modeling choices in an experimental context, our approach has implications for investigating choice behavior in the field. "Real-world" choices vary markedly in time-to-consumption, meaning experienced utility can be nearly immediate or far off. Examples abound: videos selected in-store versus those queued for future delivery from some online provider; vacation excursions selected at one's destination versus those chosen in advance at the time of purchase; projections of use of a gym membership over a year-long contract versus at the time of signing; scanner-recorded supermarket purchases made for quick consumption (e.g., produce) versus those for longer-term use (e.g., frozen vegetables); or, more generally, product classes with multiple channels of distribution that differ in time delay or—although we have not studied this directly here—in product information availability (and hence uncertainty, e.g., the possibility of taste testing in the store but not online, ability to try on clothing in a shop but not when purchasing via catalogue, etc.). An especially fertile area for future investigation is in the realm of preference elicitation methods, notably conjoint, where "purchase readiness" can mean anything from "right now" up to many months later; given our findings, it is unclear whether it is appropriate to presume that all such potential customers should be considered to have identical degrees of stochastic variation.

In sum, whenever the degree of "error"—whether it be called uncertainty, variability, noise, random component, or something else—may be more pronounced in a particular subset of the data, *no matter its source*, it need be accommodated in the formal analysis of actual choices. That is, researchers must be mindful of the fact that different conditions can entail different variances, and choose their methods accordingly. The class of models examined here offers an avenue for accommodating this distinction and examining its substantive implications.

Acknowledgments
This article is based on the first author's dissertation at the Stephen M. Ross School of Business, University of Michigan. The authors thank Christie Brown, Rich Gonzalez, Norbert Schwarz, Frank Yates, and the review team for their valuable comments. This paper will also appear in the *Festschrift to Honor John D. C. Little*.

Appendix

Given any set of deterministic components, $\{\nu_1, \ldots, \nu_K\}$, nondegenerate joint error density, $\{\varepsilon_1, \ldots, \varepsilon_K\}$, and utilities $\{\nu_j + \sigma\varepsilon_j\}$, we prove three facts: when degree of stochasticity (σ) increases, (1) the most-favored (highest v_j) item is less likely to be chosen, (2) the least-favored (smallest v_j) item is more likely to be chosen, and (3) systematic predictions cannot be made about other items. Consider the probability that item j is chosen from among a fixed set. Without loss of generality, we focus on item $j = 1$; the probability it is chosen from among items $1, \ldots, K$ is given by

$$\Pr[1] = \Pr[v_1 + \sigma\varepsilon_1 > \{v_k + \sigma\varepsilon_k\}_{k>1}]. \tag{11}$$

We can rearrange terms as follows:

$$\begin{aligned}\Pr[1] &= \Pr\left[\left\{\varepsilon_1 - \varepsilon_k > \frac{(v_k - v_1)}{\sigma}\right\}_{k>1}\right] \\ &= \Pr\left[\left\{\gamma_k > \frac{(v_k - v_1)}{\sigma}\right\}_{k>1}\right] \qquad (12)\end{aligned}$$

for error process $\{\gamma_2, \ldots, \gamma_K\}$, where $\gamma_k = \varepsilon_1 - \varepsilon_k$.

We now wish to examine the effect of changing σ. If item 1 is the most-favored alternative, $(v_k - v_1)/\sigma$ is negative for all $k > 1$. Increasing σ therefore contracts the domain of integration (over the joint pdf of $\{\gamma_2, \ldots, \gamma_K\}$) along all $K-1$ dimensions, thereby decreasing Pr[1]. An analogous argument holds for the least-favored option (when v_1 is lowest), because increasing σ expands the domain of integration. This argument does *not* hold for items other than the smallest and largest. Thus, it is established that the probability of choosing the most-favored option decreases with σ, whereas the probability of choosing the least-favored option increases for any non-degenerate error density. Note that this does not presume any particular relation amongst the errors $\{\varepsilon_k\}$, like independence, to hold.

If item 1 is neither the most nor least favored, we must show that no clear directional statements can be made about how its choice probability changes with σ; in fact, it will in general depend on the relative values of $\{v_2, \ldots, v_K\}$ (as well as the joint density $\{\varepsilon_1, \ldots, \varepsilon_K\}$). To see this, consider $K-1$ identical "favorite" items and one "nonfavorite" item. As shown above, the choice probabilities of each of the $K-1$ favorites will decrease as σ increases. Because the last expression in (12) is continuously differentiable in ν_1, this holds locally, for values of ν_1 near the favorite. (Conversely, the case of $K-1$ identical nonfavorite items and one favorite item would result in the choice probability of each of the $K-1$ nonfavorites increasing with σ.) This is readily illustrated by two examples; in both, $\{v_2, \ldots, v_K\} = \{0.98, 0.99, 1, 1.01, 1.02\}$, and for simplicity, $\{\varepsilon_1, \ldots, \varepsilon_K\}$ are i.i.d. $N[0, 1]$. If $v_1 = 2$, it is the favorite, and it is readily verified that increasing σ will decrease Pr[1] and increase *all* Pr[k], $k > 1$ (specifically, numerical integration shows that $\mathrm{d}\Pr[k]/\mathrm{d}\sigma \approx \{-0.330, 0.069, 0.067, 0.065, 0.064, 0.063\}$). If $v_1 = 0$, it is the least favorite, and increasing σ will increase Pr[1] and decrease *all* Pr[k], $k > 1$ ($\mathrm{d}\Pr[k]/\mathrm{d}\sigma \approx \{0.064, -0.007, -0.011, -0.013, -0.015, -0.019\}$). Thus, we see that although increasing σ causes the choice probabilities of the most-favored item to decrease and of the least-favored item to increase, those of the "internal" items ($k = 2, 3, 4, 5$) can be influenced positively or negatively.

References

Adamowicz, W., D. Bunch, T. A. Cameron, B. G. C. Dellaert, M. Hanneman, M. Keane, J. Louviere, R. Meyer, T. Steenburgh, J. Swait. 2008. Behavioral frontiers in choice modeling. *Marketing Lett.* **19**(3/4) 215–228.

Allenby, G. M., J. L. Ginter. 1995. The effects of in-store displays and feature advertising on consideration sets. *Internat. J. Res. Marketing* **12**(1) 67–80.

Allenby, G. M., P. E. Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econometrics* **89**(1–2) 57–78.

Anderson, C. J. 2003. The psychology of doing nothing: Forms of decision avoidance result from reason and emotion. *Psych. Bull.* **129**(1) 139–167.

Andrews, R. L., A. Ainslie, I. S. Currim. 2008. On the recoverability of choice behaviors with random coefficients choice models in the context of limited data and unobserved effects. *Management Sci.* **54**(1) 83–99.

Baltas, G., P. Doyle. 2001. Random utility models in marketing research: A survey. *J. Bus. Res.* **51**(2) 115–125.

Bhat, C. R. 1995. A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Res. B.* **29**(6) 471–483.

Bradley, M. A., A. J. Daly. 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation* **21**(2) 167–184.

Dhar, R. 1997. Consumer preference for a no-choice option. *J. Consumer Res.* **24**(2) 215–231.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica* **50**(4) 987–1008.

Harding, M. C., J. Hausman. 2006. Flexible parametric estimation of the taste distribution in random coefficients logit models. Working paper, MIT Department of Economics, Cambridge, MA.

Hensher, D. A., J. J. Louviere, J. D. Swait. 1999. Combining sources of preference data. *J. Econometrics* **89**(1–2) 197–221.

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

Jöreskog, K. G. 1994. On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* **59**(3) 381–389.

Kahneman, D., J. Snell. 1992. Predicting a changing taste: Do people know what they will like? *J. Behavioral Decision Making* **5**(3) 187–200.

Kahneman, D., P. P. Wakker, R. Sarin. 1997. Back to Bentham? Explorations of experienced utility. *Quart. J. Econom.* **112**(2) 375–405.

Keren, G., P. Roelofsma. 1995. Immediacy and certainty in intertemporal choice. *Organ. Behavioral Hum. Dec.* **63**(3) 287–297.

Kivetz, R., I. Simonson. 2002. Self-control for the righteous: Toward a theory of precommitment to indulgence. *J. Consumer Res.* **29**(2) 199–217.

Kreps, D. M. 1979. A representation theorem for "preference flexibility." *Econometrica* **47**(3) 565–577.

Lawrence, E., D. Bingham, C. Liu, V. Nair. 2008. Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics* **50**(2) 182–191.

Loewenstein, G. 2001. The creative destruction of decision research. *J. Consumer Res.* **28**(3) 499–505.

Louviere, J. J. 2001. What if consumer experiments impact variances as well as means? Response variability as a behavioral phenomenon. *J. Consumer Res.* **28**(3) 506–511.

Louviere, J. J., T. Eagle. 2006. Confound it! That pesky little scale constant messes up our convenient assumptions! *Proc. 2006 Sawtooth Software Conf., Delray Beach, FL*, Sawtooth Software, Sequim, WA, 211–228.

Louviere, J. J., R. J. Meyer. 2007. Formal choice models of informal choices: What choice modeling research can (and can't) learn from behavioral theory. N. K. Malhotra, ed. *Review of Marketing Research*, Vol. 4. M. E. Sharpe, Armonk, NY, 3–32.

Louviere, J. J., D. A. Hensher, J. D. Swait. 2000. *Stated Choice Methods*. Cambridge University Press, Cambridge, UK.

Louviere, J. J., D. Street, R. Carson, A. Ainslie, J. R. DeShazo, T. Cameron, D. Hensher, R. Kohn, T. Marley. 2002. Dissecting the random component of utility. *Marketing Lett.* **13**(3) 177–193.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.

Read, D., G. Loewenstein. 1995. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *J. Experiment. Psych.-Appl.* **1**(1) 34–49.

Read, D., G. Loewenstein, S. Kalyanaraman. 1999. Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic. *J. Behavioral Decision Making* **12**(4) 257–273.

Read, D., G. Antonides, L. van den Ouden, H. Trienekens. 2001. Which is better: Simultaneous or sequential choice? *Organ. Behav. Human Dec.* **84**(1) 54–70.

Seetharaman, P. B. 2004. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Sci.* **23**(2) 263–271.

Seetharaman, P. B., P. K. Chintagunta. 1998. A model of intertia and variety seeking with marketing variables. *Internat. J. Res. Marketing* **15**(1) 1–17.

Simonson, I. 1990. The effect of purchase quantity and timing on variety-seeking behavior. *J. Marketing Res.* **27**(2) 150–162.

Swait, J., J. Louviere. 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J. Marketing Res.* **30**(3) 305–314.

Walsh, J. W. 1995. Flexibility in consumer purchasing for uncertain future tastes. *Marketing Sci.* **14**(2) 148–165.