



Marketing Science

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

Rejoinder—Temporal Stochastic Inflation in Choice-Based Research

Linda Court Salisbury, Fred M. Feinberg,

To cite this article:

Linda Court Salisbury, Fred M. Feinberg, (2010) Rejoinder—Temporal Stochastic Inflation in Choice-Based Research. Marketing Science 29(1):32-39. <https://doi.org/10.1287/mksc.1090.0530>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2010, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Rejoinder

Temporal Stochastic Inflation in Choice-Based Research

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, feinfb@umich.edu

We examine the specification and interpretation of discrete-choice models used in behavioral theory testing, with a focus on separating “coefficient scale” from “error scale,” particularly over time. Numerous issues raised in the thoughtful commentaries of Louviere and Swait [Louviere, J., J. Swait. 2010. Discussion of “Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test.” *Marketing Sci.* 29(1) 18–22] and Hutchinson, Zauberman, and Meyer (HZM) [Hutchinson, J. W., G. Zauberman, R. Meyer. 2010. On the interpretation of temporal inflation parameters in stochastic models of judgment and choice. *Marketing Sci.* 29(1) 23–31] are addressed, specifically the roles of response scaling, preference covariates, actual versus hypothetical consumption, “immediacy,” and heterogeneity, as well as key differences between the experimental setup in Salisbury and Feinberg [Salisbury, L. C., F. M. Feinberg. 2010. Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test. *Marketing Sci.* 29(1) 1–17] and those typifying intertemporal choice and construal level theory. We strongly concur with most of the general conclusions put forth by the commentary authors, but we also emphasize a central point made in our research that may have been lost: that the temporal inflation effects observed in our empirical analysis could be attributed to stochastic effects, deterministic influences, or an amalgam; appropriate inferences depend on the nature of one’s data and stimuli. We also report on further analyses of our data, as well as a meta-analysis of HZM’s Table 1 that is consistent with our original findings. Implications for, and dimensions relevant to, future research on temporal stochastic inflation and its role in choice-based research are discussed.

Key words: brand choice; choice models; construal level theory; decisions under uncertainty; decision making over time, econometric models; intertemporal choice; measurement and inference; probability models; stochastic models

History: Received: July 3, 2009; accepted: August 4, 2009. Published online in *Articles in Advance* November 3, 2009.

Introduction

We are grateful for such insightful and thorough commentaries on our article, which examines the potential pitfalls of assuming constant stochastic variation for choices made at different times. Each commentary highlights central issues regarding choice processes, measurement, and how theories of choice behavior should be verified. Louviere and Swait (2010; hereafter L&S) focus their critique on the role of error variability in discrete-choice models, whereas Hutchinson, Zauberman, and Meyer (2010; hereafter HZM) explore implications of our article for model development and testing in a variety of research settings, including intertemporal choice (Frederick et al. 2002, Read 2004, Soman et al. 2005) and construal level theory (Troe and Liberman 2003).

We broadly agree with the commentary authors’ main points, but we also see a need to clarify certain aspects of the specification and application of the random-utility-based, discrete-choice framework that underlies our work and so much other work in decision theory, economics, and marketing. Several such points of agreement stand out. First, researchers need to apply discrete-choice methods with care and not presume (as many prior researchers have) that any observed differences arise strictly from coefficients, as opposed to error scale. Second, there is a long history of empirical findings that our work builds on, capably summarized by L&S, that merits far greater currency among behavioral researchers. Third, dedicated experiments are needed to pin down not only the true extent of temporal stochastic inflation but the roles of scaling artifacts, actual versus hypothetical

consumption, “conjoint-like” designs, and a host of other behavioral subtleties. It is this last point that much of our response addresses, illustrating key differences between our article and the many provocative studies discussed by HZM. Two such differences, which we address in the sequel, are paramount: whether HZM’s studies, taken together, challenge the claims in our article; and whether our statistical findings can be claimed to provide *unambiguous* support for temporal stochastic inflation acting alone.

We organize our discussion around three general areas: (1) discrete choice, scaling, and random utility specification; (2) psychological theories of the effect of time delay on choice; and (3) design of experimental choice tasks to test those theories. Because a comprehensive review of each area is beyond the scope of this rejoinder, we will cover each selectively, inasmuch as it pertains to the issues at hand.

Discrete Choice, Scaling, and Random Utility Specifications

Over the past two decades, Louviere, Swait, and colleagues have done the painstaking legwork of amassing a trove of evidence challenging the constant stochastic variance assumption, mainly across subjects and items (among other dimensions), but also time. We cannot do justice to this vast body of work here, but strongly encourage researchers to become familiar with it before building models that may discount extant empirical findings. In particular, L&S’s commentary contains a valuable history of the genesis of this evidence, as well as a framework for future theoretical modeling work. We would also call for renewed application of Swait and Louviere’s (1993) seminal “scaling parameter”—which all too often is assumed simply to equal one—and Louviere’s (2001) thought piece on error variance effects, which provided the spark for our decomposition into “analyst” and “consumer” error.

First, we revisit the generic utility framework referenced by L&S, HZM, and us to clarify key issues and what we view as important directions for future work in the area. The standard utility-based representation underlying choice models includes three components: covariates, coefficients, and error. Each of these can differ across individuals (i), alternatives/brands (j), and time (t):

$$U_{ijt} = X_{ijt}\beta_{ijt} + \varepsilon_{ijt}.$$

This representation for U_{ijt} can, of course, hold for ratings (on a fixed scale) as well as for discrete choices, an issue to which we will return. There is complete agreement in the literature that covariates, X_{ijt} , can and do differ across all three dimensions. For the behavioral researcher and econometrician, there is less agreement when we consider whether coefficients

(β_{ijt}) and error (ε_{ijt} , to be later represented by its standard deviation) can change across any (or all) individuals, alternatives, and time.

Early discrete-choice models—particularly the McFadden (1974) conditional logit ubiquitous in classic empirical work in marketing—typically imposed restrictive assumptions on β_{ijt} and ε_{ijt} to lend stability in estimation. Specifically, β_{ijt} was decomposed into brand-specific constants (which differ across brands, but not individuals or time) and other coefficients, which were taken to be identical across i , j , and t . Moreover, error variance was typically presumed constant across all three dimensions.

Bayesian analysis and fast computing enabled a far wider variety of coefficient specifications. It is generally trivial to allow coefficients to differ across alternatives (j), so long as they are modest in number. It is also possible to allow for unobserved heterogeneity, that is, for coefficients to differ across individuals (i), even though each made very few choices.¹ Recent work in choice dynamics has modeled variations in coefficients across time (t) as well (e.g., Kim et al. 2005, Lachaab et al. 2006). Similar generalizations are possible for errors, which can vary: across alternatives (j) in a “logit-like” heteroscedastic extreme value model or in a more general multinomial probit specification; across individuals (i) using the “generalized” multinomial logit model (Fiebig et al. 2009, Feit 2009) or latent classes with Swait-Louviere scale factors (Kanetkar and Islam 2009); and even, to some degree, across time (t)—for example, using an autoregressive process (e.g., Keane 1997, Lachaab et al. 2006). The focal model in our paper allowed error scale to vary across alternatives (j) and across time (t).

The key issue in our work and the two commentaries involves the separate identification of error scale and coefficients. The work of Fiebig et al. (2009) and Feit (2009) shows that (unobserved) heterogeneity (across individuals, i) in both coefficients and error scale can be *jointly* specified, recovered reliably in simulated data, and measured in real data sets; the latter paper summarizes prior work on the covariates shown to affect error scale and “choice consistency,” including price differences, fatigue, attribute levels, and response times, among others. These papers challenge the long-held notion that in discrete-choice models, one could not separately identify (across individuals, i) both coefficients and error variation, but only their ratio. The question remains about whether

¹ However, a stream of recent research suggests caution even here. Sonnier et al. (2007) show that imposing heterogeneity on coefficients for latent utilities can entail substantively different guidance than imposing it on other, “meaningful” quantities, like willingness to pay. Andrews et al. (2008) show that, with small numbers of choices, individual parameters may not be accurately captured even when the model does recover the true data-generating process.

this is plausible across time (t). We agree wholeheartedly with both commentaries that, given our results, one cannot attribute measured “temporal inflation” *wholly and unambiguously* to error scale differences across temporal conditions. Our main point is that it is just as unfounded to attribute it—as many prior analyses have done—wholly and unambiguously to differences in *coefficients*. We would go further than HZM, however, to suggest that, in carefully designed experiments, one could justifiably attribute such inflation primarily to error scale differences. A critical point is that the nature of the choice task examined dictates which assumptions are reasonable to make.

Early on, HZM (p. 24) make an essential remark, which we reproduce in full: “Interpretation of the scale parameter depends on what assumptions one is willing to make Salisbury and Feinberg (2009) assume that the *temporal inflation* parameter σ_c represents differences in internal noise between temporally near versus distant predictions about experienced preferences.” We agree strongly with the first part but feel a clarifying comment on the second is appropriate. Specifically, not all assumptions are created equal: as we will contend subsequently, assumptions reasonable in the sorts of choice task examined in HZM’s commentary are far less so in the choice task examined in our article. Furthermore, we would make a critical amendment, differentiating a statement and its converse: rather than assuming that σ_c solely represents stochastic differences between temporally near and far, we warn instead against presuming that σ_c *cannot* represent such differences. This is not merely a matter of phraseology or philosophy: essentially all prior work *has* made just such assumptions. The purpose of our paper is to examine the implications of these presumptions—not to supplant them with equally questionable ones. In the language of HZM’s models 1 and 2, we are arguing that interpreting σ_c as unambiguously supporting either purely deterministic or purely stochastic effects is premature and can lead to faulty substantive conclusions.

Testing Theories of Time Delay and Choice

HZM discuss intertemporal choice (ITC) and construal level theory (CLT) at some length. Our theoretical framework and experimental choice task differ from both in important ways, as summarized in Table 1 (which we discuss later). It is helpful to recall generic definitions of these choice domains and theories; first, (*italics our own*):

‘Intertemporal choice’ is used to describe any decision that requires *tradeoffs among outcomes* that will

have their *effects at different times* (Read 2004, p. 424); Intertemporal choice refers to a *choice between options* whose *consequences occur at different points in time*. (Soman et al. 2005, p. 348)

It should be clear that, by these definitions, our work has little to do with classic intertemporal choice: trade-offs are never enacted, either within- or between-subjects, across time. In our experiments, which echo those of Simonson (1990) and Read and Loewenstein (1995), the key manipulated quantity is *time between choice and consumption*. In the classic intertemporal choice task, the decision maker is presumed to trade off time to consumption with other attributes of the available alternatives (e.g., choose between receiving one cookie today versus two cookies next week). In contrast, our choice task includes alternatives whose *outcomes will all be experienced at the same time* (e.g., choose between receiving a cookie next week or a candy bar next week); no intertemporal trade-offs are required. Similarly, “discounting,” as classically defined, plays no explicit role in our decision task; our two conditions involve identical consumption schedules. All that comes into play is when subjects choose what they will consume within those predetermined schedules. (We will return to the issue of consumption later.)

CLT research examines the effects of psychological distance on mental representations (or construals) of available alternatives. This includes temporal distance, spatial distance, hypotheticality, and social distance from an object or event (see Liberman et al. 2007 for a detailed discussion). A central tenet of CLT is summarized by Trope et al. (2007, p. 83):

CLT assumes that people mentally construe objects that are psychologically near in terms of low-level, detailed, and contextualized features, whereas at a distance they construe the same objects or events in terms of high-level, abstract, and stable characteristics [T]hese construals, in turn, guide prediction, evaluation, and behavior.

Within the context of time delay and choice, “CLT proposes that people use higher-level construals for distant future events than for near future events. It therefore predicts that the value associated with low-level construals should be more prominent in a subjective evaluation of near future events, whereas the value associated with high-level construals should be more prominent in evaluating distant future events.” (Liberman and Trope 2003, pp. 249–250). More specifically, CLT studies examine the effect of time delay on the relative weighting of attributes or features of the alternatives under consideration. Tests of CLT, therefore, require that various product features be devised and manipulated to be differentially salient at various times. Our experiment deliberately avoids this

construct, with substantial implications for variance inflation, as we discuss below.

The typical choice task employed by CLT researchers (examining temporal construal effects) resembles ours in that the available alternatives entail outcomes occurring at the same time. However, unlike our choice task, the stimuli typically involve hypothetical alternatives and include researcher-defined attributes and attribute levels, and preference is often inferred from relative rating scores. For example, experiments discussed in HZM ask subjects to allocate 100 points between two options (Malkoc et al. 2005), rate liking and satisfaction (Trope and Liberman 2000), indicate relative preference using a scale (1 = definitely option A and 10 = definitely option B; Zhao et al. 2007), or state likelihood of choosing (Zhao et al. 2007). By contrast, Malkoc et al. (2005, experiment 2) and Zhao et al. (2007, experiment 2) do ask subjects to make a choice, albeit binary. To our knowledge, none of the CLT work cited involved multinomial choice, as our task did, although this in itself is likely an ancillary issue.

Also, to the best of our knowledge, there has been no published attempt to formalize CLT mathematically. HZM do so in Equations (7a) and (7b) for their specific application of CLT to the effects of temporal distance on relative weighting of product attributes. This is a welcome advance in the application of CLT to understand the effects of time delay on choice behavior, and further research would do well to develop models that more comprehensively generalize CLT in other contexts. Dedicated experiments designed specifically to test whether temporal inflation is supported in ratings data will require that the assumption clearly (and, in our view, correctly) stated as not being universally true by HZM (p. 26)—“when it is reasonable to assume that the same evaluation process is used in both tasks and only the response component differs”—is rigorously tested.

In Table 1, we lay out a number of important dimensions in the theory and experimental operationalizations of both ITC and CLT, and how they each contrast with deliberately selected elements of our approach. Broadly speaking, ITC differs in its theory and the nature of trade-offs subjects are asked to make, and CLT differs markedly in its operationalizations. Because the core construct in ITC—trade-offs in time itself—plays no role in our work, we will focus mainly on CLT, as do HZM.

Experimental Choice Task Design

HZM have ably summarized several experiments that may call certain aspects of temporal stochastic inflation into question. They present evidence that, in a variety of tasks primarily centered on CLT or

time discounting (of consumption costs), the “temporal inflation factor” on error standard deviation—the focus of our paper—does not appear to be greater than one. We have no quarrel with these findings, as they are, for reasons we present now, largely what we would have predicted ourselves. Although HZM’s interpretation is entirely appropriate for their specific class of research tasks, a rather large number of caveats need to be considered before the general matter of temporal stochastic inflation is settled. Because this cuts to the heart of the matter, it is helpful to examine these in some detail.

HZM’s work in this area is valuable not only for its lucid presentation of CLT but for helping to delineate what we see as a $2 \times 2 \times 2$ research design “cube” for further progress in the area, with their work occupying the corner diametric to ours. The dimensions of this cube, on which we will expound at some length, can be summarized as follows:

1. *Scales*: Imposed versus Latent,
2. *Stimuli and Preference Covariates*: Artificial/Defined versus Actual/Self-stated, and
3. *Consumption and IMM*: Hypothetical/Delayed versus Actual/Immediate.

The number of differences falling under these broad headings is nontrivial; thus we offer brief discussions of each as a way to frame possibilities for future research that could “fill in the cube” and thereby shed light on the existence and extent of temporal stochastic inflation.

Without opening up an ancient and woolly debate in the philosophy of science, one might say that the research cited by HZM opts more for “internal validity”—using imposed scales, artificial (but precisely defined) stimuli, and hypothetical consumption—whereas we opt more for “external validity,” allowing subjects to choose among real, well-known products and then to actually consume their choices in accordance with the tenets of the theory we seek to examine. For reasons we will detail shortly, none of this is desirable, or perhaps even possible, in direct tests of CLT, the domain into which the strong majority of the research quoted by HZM falls.

Scales: Imposed vs. Latent

Rating scales have been workhorses in behavioral research for decades, and their psychometric properties and strengths are well understood. Here, we discuss their role in a highly specific context: relative to choice-based elicitation methods, particularly when the critical comparisons involve variances, as they do in our research. For example, conjoint research has largely moved away from imposed rating scales, for a variety of reasons, chiefly because they are less “natural” for participants. Ample evidence suggests that subjects should do what is simple for them (choosing) and the analyst should do the work of scaling,

Table 1 A Comparison of the Three Choice Contexts

S&F choice task	Classic intertemporal choice task ^a	Typical (temporal) CLT task ^a
<i>Objective:</i> examine effect of time delay on stochastic variance in discrete choice.	<i>Objective:</i> examine effect of time delay on relative attractiveness of alternatives whose outcomes are experienced at different times.	<i>Objective:</i> examine effect of time delay on relative weighting of attributes.
<i>Task:</i> Encompasses multiple, $n > 1$, choices: {A now; B now} and {A later; B later} and {A later; B even later}.	<i>Task:</i> Encompasses a single, $n = 1$, choice: {A now; B later}.	<i>Task:</i> Encompasses a single, $n = 1$, choice: {A later; B later}.
Alternatives do <i>not</i> differ in outcome time, t .	Alternatives differ in outcome time, t .	Alternatives do <i>not</i> differ in outcome time, t .
Consumption occurs at time t .	Consumption occurs at time t or later, and is often hypothetical.	Consumption is often hypothetical.
No attribute information supplied.	Outcome time is an attribute.	Attribute types and levels supplied.
DV: choice	DV: choice, willingness to pay	DV: rating scale, sometimes choice
“Mixed” design	Within-subjects design	Between-subjects design
$U_{ijt} = V_{ijt} + \theta_c \sigma_c \xi_{ijt}$ $V_{ijt} = f(\text{RATE}_{ijt})$.	$U_{ijt} = V_{ijt} / \sigma_c + \varepsilon_{ijt}$ $V_{ijt} = f(\text{assumed attractiveness}, \sigma_c)$.	$U_{ijt} = V_{ijt} / \sigma_{ck} + \varepsilon_{ijt}$ $V_{ijt} = f(\text{ATTRIBUTE}_{jk}, \sigma_{ck})$.
Consumers do <i>not</i> make a trade-off between outcome time and some other attribute(s). No “intertemporal trade-offs” made.	Consumers make a trade-off between outcome time and attractiveness. “Intertemporal trade-offs” made.	Consumers do <i>not</i> make a trade-off between outcome time and some other attribute(s). No “intertemporal trade-offs” made. Other attribute trade-offs are made.
<i>Key hypothesis:</i> σ_c is larger for choice $t > 0$ than for choice $t = 0$ $\sigma_{t>0} > \sigma_{t=0}$	<i>Key hypothesis:</i> σ_c is larger for alternative B than for A $\sigma_{j=B} > \sigma_{j=A}$	<i>Key hypothesis:</i> ^b σ_{ck} is larger/smaller for choice $t > 0$ versus choice $t = 0$ $(\sigma_{t>0, l} / \sigma_{t=0, l}) > (\sigma_{t>0, h} / \sigma_{t=0, h})$
All alternatives are “inflated” for choices $t > 0$.	Only alternative B is “inflated” (discounted).	Attribute weights inflate or deflate for $t > 0$.
σ_c is the same for all alternatives.	σ_c is <i>not</i> the same for all alternatives.	σ_{ck} is the same for all alternatives.
<i>Prediction</i> (assuming B has greater mean attractiveness): $\Pr(\text{choose A}, t > 0) > \Pr(\text{choose A}, t = 0)$.	<i>Prediction</i> (assuming B has greater mean attractiveness): $\Pr(\text{choose A}) > \Pr(\text{choose B})$.	<i>Prediction</i> (assuming A dominates on h , and B dominates on l): $\Pr(\text{choose A}, t > 0) > \Pr(\text{choose A}, t = 0)$, $\Pr(\text{choose B}, t > 0) < \Pr(\text{choose B}, t = 0)$.

^aWe adopt notation consistent with HZM’s models 2 and 3 as a basis for comparison only. It is not intended to reflect other extant ITC or CLT literature.

^bWe denote “high-level” attributes as h and “low-level” attributes as l .

rather than forcing subjects to place their answers on an artificial scale, no matter how broad or granular. For concision and clarity, it is difficult to surpass Lenk and Bacon’s (2008, p. 1) opening remarks on the topic:

Discrete-choice elicitation is often preferred to other measurement methods because it better aligns with actual choice behavior and avoids some of the well-documented biases inherent to alternative methods, such as ratings. A limitation of discrete-choice is the loss of inter-subject comparability because the resulting utility structure is invariant to linear transformations.

Their focus was on between-subjects comparisons, but the same issues bedevil those across time.

For example, research in psychometrics (e.g., Lenk et al. 2006), cross-cultural comparison (e.g., Yoon et al. 2004) and marketing proper (e.g., Rossi et al. 2001) suggests individuals use scales idiosyncratically. Although broadly validated in testing for mean

differences, it is far from established that standard ratings scales are stable, homogeneously utilized, and reliable enough to enable their use in comparing variances across experimental settings, nor, in fact, to form the basis of comparison of the observed scale-based variance ratio and the latent temporal inflation factor. In simple terms, it is entirely possible that subjects in each experimental group use the scale to allow maximal discriminability across the set of questions they are, or expect to be, asked, so that any between-group latent variance differences would be strongly attenuated, as HZM found.

Stimuli and Preference Covariates:

Artificial/Defined vs. Actual/Self-Stated

Use of (Homogeneous) Imposed Attributes and Levels. In all the research quoted by HZM, and much besides, the focal task might be called “conjoint-like”: alternatives are literally *defined* by their attributes. As

such, the potential is rife for subjects' own views of what is critical in the product class to be overridden by what they are presented by the researcher. To define alternatives by attributes requires that values (levels or part worths) themselves must be provided. In such cases, and unlike in real product choice contexts, a one-unit difference (in two products' utilities) on "attribute 1" is unambiguously smaller—in fact, precisely half—than a two-unit difference elsewhere. The "cost of thinking" (Shugan 1980) in making such comparisons is small and nearly zero when one product literally dominates another. This serves to render a particular type of "uncertainty"—what we posit to underlie the very effect we are studying—dramatically less relevant. Because our choice task used well-known brands, and so lacked given attributes or levels, this potential confound was avoided. Finally, there is *heterogeneity*: when attributes and levels are determined by the researcher, heterogeneity can only arise from subjects' different valuation of the given attributes, not from the levels themselves (i.e., disagreement over how well products score on an attribute) nor from nonincluded attributes. Both would, again, attenuate uncertainty. By contrast, our task accounted for a specific type of observed heterogeneity, via self-reported ratings, to which we turn next.

Self-Stated Summary vs. Imposed Attribute-Based Preference Measures. Our task and utility model made use of synthetic, individual-level measures of overall product attractiveness, *RATE*. Subjects could thereby bring to bear their own bases for comparison among the alternatives (attributes), evaluate each (levels), and weight them idiosyncratically. This measure was collected a week in advance of the first experimental session, to minimize the potential for demand artifacts at time of choice. If "utility" possesses anything other than a purely latent meaning, it is difficult to see, a priori, why a synthetic, individualized measure of quality would be systematically downplayed in a choice made for one week, versus two weeks, from now (as might be consistent with classical discounting or the scaling of HZM's model 2). This observation is well in line with HZM's (accurate) remark that interpreting identification in discrete choice models must be resolved by assumption, with the caveat that some assumptions are dramatically more plausible than others, depending on experimental design and context.

Consumption and IMM: Hypothetical/Delayed vs. Actual/Immediate

A critical concern in our research is the time between choice and consumption. As such, we manipulated this quantity directly: participants consumed their chosen alternative immediately (*IMM*) after

choice, before completing the experimental session; or participants consumed the chosen alternative during a future (*FUT*) prescheduled experimental session occurring one or two weeks out. This contrasts sharply with the manipulations used in the experiments cited by HZM: *IMM* conditions involved time delays of up to a week, and none of the CLT tests included actual consumption during the experimental session.² In other words, our *IMM* condition used actual/immediate consumption, whereas the others used hypothetical/delayed consumption. This is a crucial difference that, we would contend, affects our core construct, temporal stochastic inflation. Choosing what to consume immediately (i.e., here and now) would likely invoke different levels of choice uncertainty than "near" consumption that could be up to a week away (as in two of the experiments reported by HZM). For our choice domain, *IMM* has a very precise meaning, whereas for CLT, *IMM* is a special case of "near." There is also the key issue of whether people would choose differently for a real choice that they have committed to consuming on site versus a hypothetical choice that they may expect to never consume.

We wish to emphasize that the methods and measures used in the studies of HZM's Table 1 are entirely appropriate to their original contexts. In our view, however, there is one potentially insurmountable difficulty in using results from CLT studies to test temporal inflation: the attributes (and levels) selected for such studies were (appropriately) devised to be differentially salient in temporally near (*IMM*) versus distant (*FUT*) choice contexts. For example, HZM (p. 27) note that the experimental stimuli used by Malkoc et al. (2005, experiment 1) were "designed so that one product was more favorable on shared attributes . . . and the other . . . on unique attributes." Given the use of identical scales, different groups of respondents, stimuli created to "engage" at different times, and without the actual passage of time (as in our study), it is not entirely surprising that scale usage variation fails to differ significantly across groups (that is, as a proxy for time). It is for these reasons that we cannot concur with HZM's contention that "this shift in preference without a shift in variance supports models 2 and 3" (p. 27), unless one can be sure that the studies in question were free of between-subjects scale-usage artifacts and experimental manipulations (appropriately, in their original contexts) intended to reduce variation between conditions.

² One potential exception is experiment of Malkoc et al. (2005), in which "Participants in the near-future condition received their selected brand of potato chips after approximately 1 hr, and participants in the distant-future condition received their preferred brand after 3 weeks" (p. 414).

Additional Empirical Evidence

Meta-Analysis of HZM's Table 1. The main data brought to bear by HZM are contained in their Table 1 (hereafter, HZM-Table 1). For reasons already discussed, the jury is still out on whether data such as in their last column (σ_{FUT}) speak to the existence of temporal variance inflation. However, HZM-Table 1 includes another sort of data, on whether choice probabilities “move toward” 0.5 (or, more generally, $1/n$) from *IMM* to *FUT*. These derive from four papers³ and six separate experiments (MZU1, MZU2, ZL, ZHZ1, ZHZ2, Z), with sample sizes of 177, 151, 264, 189, 225, and 70, respectively; there are conditions within these experiments as well, bringing the total number of “proportion pairs” (e.g., 0.21 versus 0.30 for MZU1) reported in HZM-Table 1 to 12. The studies giving rise to HZM-Table 1 were carefully conducted and, as its associated footnote indicates, their differential discounting model makes a set of predictions indicated (by + and – signs) in the Choice_{IMM} and Choice_{FUT} columns.

We perform a meta-analysis on these data to determine whether proportions indeed “move toward” 0.5, which is the critical comparison for temporal stochastic inflation. By coding each proportion between 0 and 0.5 (i.e., subtracting from 1, if necessary), we find the weighted (by sample size) mean proportions to be 31.8% for *IMM* and 35.3% for *FUT*, a 3.5% absolute, and 10.9% proportional, difference. Testing this toward-0.5 motion is readily accomplished via logit transforms and nonlinear optimization. With 1,076 total subjects, this yields $t = 12.7$, which is very strong evidence of the “toward $1/n$ shift” we hypothesized and which one might not see in the disaggregated results of HZM-Table 1.⁴ Thus, we cannot agree with HZM's contention that the results of HZM-Table 1 fail to support our analysis. In fact, given that the “toward $1/n$ shift” is a robust, nonparametric criterion, we believe HZM-Table 1 provides unusually compelling evidence for our basic finding. It does not, however, help distinguish between models 1 and 2, and we agree with HZM that further studies need to be specifically designed to that end.

Does the Inflation Factor Vary Over Time? HZM point out that “Salisbury and Feinberg report that estimating separate inflation factors for $t = 2$ and $t = 3$ does not significantly improve the model. This

is consistent with the nearly universal finding of so-called ‘hyperbolic’ discounting” (p. 26). We did indeed find some difference in inflation between periods 2 and 3, although it was modest. Separately estimated, $\sigma_{c,2} = 2.140$ and $\sigma_{c,3} = 2.664$, $p > 0.5$ (the jointly estimated value was 2.37).⁵ However, these estimated values do “look” different, and their lack of significance may be a result of the relatively small sample sizes involved. Although discounting plays no explicit role in our work, we point out that the hyperbolic variety disproportionately erodes the distinction between “now” and “later” when a fixed time interval is added to both. Because its *IMM* conditions were not, unlike ours, truly immediate (in terms of consumption), the final column of HZM-Table 1 is very much in line with hyperbolic discounting. Given the absence of actual consumption, calculating a discount rate, or function consistent with their reported data, is daunting.

Parameters and Inference Making

We would like to stress a point made clearly by both L&S and HZM—that it is critical to avoid attributing psychological effects or conclusions about behavioral processes wholly to either deterministic or stochastic components. Rather, the best one can say is that the measured value of σ_c must be viewed as “distributed across” deterministic and stochastic explanations; the actual apportionment can be affected by experimental design, the nature of observables, and of course, what assumptions researchers enact. We also agree with HZM that “uncertainty” is different from “error” (p. 23); in fact, a key part of our analysis was to provide a decomposition of the stochastic component in discrete-choice models along the lines of “analyst” and “consumer” error (Louviere 2001). Finally, we cannot agree more strongly with the points made by HZM (p. 23), that “the effects ‘can’ be due to stochastic factors; these explanations have not been ruled out by most analyses reported in the literature” and also that “less careful readers might wrongly conclude that the scale parameter must always reflect stochastic error” (p. 24). That is, our findings are about “can,” not “must.”

In closing, we especially wish to emphasize HZM's (p. 28) entreaty that “experiments specifically designed to tease apart the various models (including hybrid combinations) will be needed to settle the issue.” Steps are already being taken in this direction. For example, Lenk and Bacon (2008) take a sizable theoretical and empirical stride by using auxiliary

³ MZU refers to Malkoc et al. (2005); ZL refers to Zauberman and Lynch (2005); ZHZ refers to Zhao et al. (2007); Z refers to Zauberman (2003).

⁴ There are many ways to run this analysis, but all appear to strongly support the shift; even the far weaker—as it ignores variation across the individual studies—and less appropriate standard classical test for the difference in the 0.353 *FUT* versus 0.318 *IMM* proportions yields $t = 2.42$, $p < 0.01$.

⁵ HZM also inquire as to the “ABA pattern” proportion—which Read and Loewenstein (1995) found to be far larger in simultaneous (55%–75%) than in sequential (21%–35%) choice—for subjects who chose two distinct items. Although the pattern is similar for our data, 12/17 (SIM) versus 13/28 (SEQ), the difference is non-significant (Fisher's exact test: $p = 0.135$.) However, sample sizes are small.

data to recover the origin and scale in discrete-choice applications, using a full hierarchical Bayes set-up, a method previously found useful in so-called MaxDiff scaling applications (Marley and Louviere 2005). We enthusiastically support the calls from both L&S and HZM for greater interaction among behavioral and quantitative researchers to advance the field in this area. HZM (p. 28) note that they “have no quarrel with... the idea that the effects of experimental variables on unobserved variance can distort estimated parameters in choice analysis.” They go on, however, to express concern that not all readers of our work will stop there, a concern we share. Another concern we share is that terms like “preference,” “uncertainty,” “utility,” “variation,” and, most notably from our perspective, “error” appear in manifold guises throughout the decision literature. Rectifying the role of “error” scale provides an opportunity to help standardize the uses of these terms throughout choice theory.

In the end, our main point is a simple one: interpreting “parameters” in choice models cannot presume they have nothing to do with the so-called “error.” We hope that our work, along with the mass of further evidence compiled by Louviere and Swait, will spur greater examination of tacit assumptions underlying commonly applied choice models.

References

- 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.
- Feit, E. M. 2009. Product category expertise and logit error scale. Working paper, Ross School of Business, University of Michigan, Ann Arbor.
- Fiebig, D. G., M. P. Keane, J. J. Louviere, N. Wasi. 2009. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Sci.*, ePub ahead of print July 23, <http://mktsci.journal.informs.org/cgi/content/abstract/mksc.1090.0508v1>.
- Frederick, S., G. Loewenstein, T. O'Donoghue. 2002. Time discounting and time preference: A critical review. *J. Econom. Literature* **40**(2) 351–401.
- Hutchinson, J. W., G. Zauberman, R. Meyer. 2010. On the interpretation of temporal inflation parameters in stochastic models of judgment and choice. *Marketing Sci.* **29**(1) 23–31.
- Kanetkar, V., T. Islam. 2009. Latent segments or scale variations: A simple choice model to incorporate heterogeneity. Working paper, University of Guelph, Guelph, ON, Canada.
- Keane, M. P. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econ. Stat.* **15**(3) 310–327.
- Kim, J. G., U. Menzefricke, F. M. Feinberg. 2005. Modeling parametric evolution in a random utility framework. *J. Bus. Econom. Statist.* **23**(3) 282–294.
- Lachaab, M., A. Ansari, K. Jedidi, A. Trabelsi. 2006. Modeling preference evolution in discrete choice models: A Bayesian state-space approach. *Quant. Marketing Econom.* **4**(1) 57–81.
- Lenk, P., L. Bacon. 2008. Estimating common utility origins and scales in discrete-choice conjoint with auxiliary data. Working paper, Ross School of Business, University of Michigan, Ann Arbor.
- Lenk, P., M. Wedel, U. Bockenholt. 2006. Bayesian estimation of circumplex models subject to prior theory constraints and scale-usage bias. *Psychometrika* **71**(1) 33–55.
- Liberman, N., Y. Trope. 2003. Construal level theory of intertemporal judgment and decision. G. Loewenstein, D. Read, R. F. Baumeister, eds. *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. Sage, New York, 245–276.
- Liberman, N., Y. Trope, E. Stephan. 2007. Psychological distance. A. W. Kruglanski, E. T. Higgins, eds. *Social Psychology: Handbook of Basic Principles*. Guilford Press, New York, 353–381.
- 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. Swait. 2010. Discussion of “Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test.” *Marketing Sci.* **29**(1) 18–22.
- Malkoc, S. A., G. Zauberman, C. Ulu. 2005. Consuming now or later? The interactive effect of timing and attribute alignability. *Psychol. Sci.* **16**(5) 411–417.
- Marley, A. A. J., J. J. Louviere. 2005. Some probabilistic models of best, worst, and best-worst choices. *J. Math. Psych.* **49**(6) 464–480.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, New York, 105–142.
- Read, D. 2004. Intertemporal choice. D. J. Koehler, N. Harvey, eds. *Blackwell Handbook of Judgment and Decision Making*. Oxford, Blackwell Publishing, Oxford, UK, 424–443.
- Read, D., G. Loewenstein. 1995. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *J. Exp. Psychol. Appl.* **1**(1) 34–49.
- Rossi, P. E., Z. Gilula, G. M. Allenby. 2001. Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statist. Assoc.* **96**(March) 20–31.
- Salisbury, L. C., F. M. Feinberg. 2010. Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test. *Marketing Sci.* **29**(1) 1–17.
- Shugan, S. M. 1980. The cost of thinking. *J. Consumer Res.* **7**(2) 99–111.
- Simonson, I. 1990. The effect of purchase quantity and timing on variety-seeking behavior. *J. Marketing Res.* **27**(2) 150–162.
- Soman, D., G. Ainslie, S. Frederick, X. Li, J. Lynch, P. Moreau, A. Mitchell et al. 2005. The psychology of intertemporal discounting: Why are distant events valued differently from proximal ones? *Marketing Lett.* **16**(3/4) 347–360.
- Sonnier, G., A. Ainslie, T. Otter. 2007. Heterogeneity distributions of willingness-to-pay in choice models. *Quant. Marketing Econ.* **5**(3) 313–331.
- 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.
- Trope, Y., N. Liberman. 2000. Temporal construal and time-dependent changes in preference. *J. Pers. Soc. Psychol.* **79**(6) 876–889.
- Trope, Y., N. Liberman. 2003. Temporal construal. *Psychol. Rev.* **110**(3) 403–421.
- Trope, Y., N. Liberman, C. Wakslak. 2007. Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *J. Consumer Psychol.* **17**(2) 83–95.
- Yoon, C., F. M. Feinberg, P. Hu, A. H. Gutches, T. Hedden, H. Y. Chen, Q. Jing, Y. Cui, D. C. Park. 2004. Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese adults. *Psych. Aging* **19**(3) 379–393.
- Zhao, M., S. Hoeffler, G. Zauberman. 2007. Mental simulation and preference consistency over time: The role of process- versus outcome-focused thoughts. *J. Marketing Res.* **44**(3) 379–388.