

## Commentary

On the Interpretation of Temporal Inflation Parameters  
in Stochastic Models of Judgment and Choice

J. Wesley Hutchinson, Gal Zauberman, Robert Meyer

The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104  
{jwhutch@wharton.upenn.edu, zauberman@wharton.upenn.edu, meyrerr@wharton.upenn.edu}

The implications of Salisbury and Feinberg's (2009) paper [Salisbury, L. C., F. M. Feinberg, 2009. Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test. *Marketing Sci.*, ePub ahead of print November 3, <http://mktsci.journal.informs.org/cgi/content/abstract/mksc.1080.0464v1>] for the process of model development and testing in the field of intertemporal choice analysis is explored. Although supporting the overall thrust of Salisbury and Feinberg's critique of previous empirical work in the area, we also see their paper as illustrating the dangers of drawing strong inferences about the behavioral interpretation of statistical model parameters without seeking convergent empirical evidence. In particular, we are skeptical about the extent to which the reported effects of temporal distance on the estimated scale parameter,  $\sigma_c$ , are uniquely, or even primarily, due to unobserved error inflation that reflects consumer's uncertainty about future utility. This interpretation is brought into question by several lines of reasoning. Conceptually, we note that "uncertainty" is different from "error" and that, for choice data, the error inflation model is mathematically identical to a model in which the scale parameter is a deterministic function of the temporal discount rate. Empirically, a reanalysis of data from previously published experiments does not consistently support temporal error inflation, temporal convergence of choice shares, or the scale parameter as an explanation of variety seeking in choice sequences. In our opinion, the cumulative results of research on intertemporal choice require models in which the attributes of choice alternatives are differentially discounted over time. Despite these findings, we advocate that choice researchers should indeed follow Salisbury and Feinberg's advice to not assume that error variances will be unaffected by experimental manipulations, and such effects should be explicitly modeled. We also agree that uncovering effects on error variance is just the first step, and the ultimate goal should be to rigorously explain the *reasons* for such effects.

*Key words:* behavioral decision theory; choice modeling; measurement and inference; behavioral economics; random utility models; intertemporal choice; psychological process models

*History:* Received: March 26, 2009; accepted: April 1, 2009. Published online in *Articles in Advance*.

## Introduction

Salisbury and Feinberg (2009) provide an important example of how the stochastic modeling of data from behavioral experiments can (1) provide insights that are not possible from standard analysis of variance models of data from experiments, and (2) provide tests of the theories underlying choice models not possible for observational stated and revealed preference data that lack experimental manipulations. We strongly agree with their general conclusion that invalid conclusions can result when the effects of experimental treatments on unobserved variability are ignored. In the simplest case, the effects of "mere" sampling error can create apparent effects on observed means. Famous examples can be found in calibration research on overconfidence (Brenner 2000, Erev et al. 1994, Wallsten et al. 2000; also see the September 1997 issue of the *Journal of Behavioral Decision Making* and Lichtenstein et al. 1982) and statistical critiques of

meta-analysis (Hunter and Schmidt 2004). The important criticism in these cases is not that the effects reported in the literature "must" be due to stochastic rather than deterministic factors, but that the effects "can" be due to stochastic factors; these explanations have not been ruled out by most analyses reported in the literature. In consumer research, Hutchinson et al. (2000) show that unobserved variation *across* individuals can produce differences between aggregate means that do not accurately represent individual-level behavior. In their current work, Salisbury and Feinberg (2009) extend a set of ideas first introduced in an earlier article (Salisbury and Feinberg 2008) showing how unobserved variation *within* individuals can produce differences between estimated coefficients (which are implicitly aggregate means) that do not accurately represent individual-level behavior.

Although we agree with Salisbury and Feinberg's (2009) general argument that behavioral researchers

should pay closer attention to the potentially confounding effects of experimental manipulations on unobserved variance when drawing conclusions about behavioral processes (see also Louviere 2001), we also see the paper as illustrating the risks of going too far in the other direction—that of drawing overly strong behavioral inferences about process from estimated model parameters, especially when those parameters are open to multiple interpretations. Actually, Salisbury and Feinberg are commendably careful in their claims; however, this care is likely to be overlooked by readers. For example, they make the fairly weak claim that if it is assumed that the time of consumption (immediate versus future) has no deterministic effect, then the estimated scale parameter,  $\sigma_c$ , of a random utility model can be interpreted as an effect on stochastic error. They are not as explicit about the converse: if it is assumed that time of consumption has no effect on stochastic error, then the estimated scale parameter can be interpreted as a deterministic effect. Less careful readers might wrongly conclude that the scale parameter must always reflect stochastic error.

To elaborate, the authors interpret the scale parameter as representing the decision-maker's "uncertainty about anticipated utility." In this commentary we argue that this interpretation is questionable for five reasons. First, error variance parameters in a stochastic model of individual behavior are best interpreted as representing the "internal noise" of the perceptual and cognitive processes that generate observed behaviors. In contrast, uncertainty can affect the decisions of rational agents who, in theory, have no internal noise in their decision-making process and thus make the same optimal decision whenever presented with the same choice alternatives. Second, when only choice data are analyzed, the scale parameter is fundamentally indeterminate with respect to whether it represents a deterministic or stochastic component. Interpretation of the scale parameter depends on what assumptions one is willing to make (as noted earlier). Salisbury and Feinberg (2009) assume that the *temporal inflation* parameter  $\sigma_c$  represents differences in internal noise between temporally near versus distant predictions about experienced preferences. We challenge this assumption with an alternative interpretation of  $\sigma_c$  as being a deterministic function of the discount rate that is used in traditional economic theories of valuing future consumption. Third, we argue that although the scale parameter is fundamentally ambiguous when only choice data are analyzed, this is not always true for judgment data. We reanalyze previously published simultaneous choice and judgment data, and find that tests of temporal inflation based on estimated error variance in preference ratings do not support the

internal noise interpretation of  $\sigma_c$ . Fourth, we note that previously reported analyses of specific choice sequences support an effect of variety seeking that is not consistent with temporal inflation. Finally, we argue that both internal noise and discounted utility of consumption are insufficient explanations of several robust phenomena observed in intertemporal choice experiments and propose a third model that we believe better represents recent work in this area.

## Deterministic and Stochastic Components of Uncertainty

In his presidential address to the Psychometric Society, R. Duncan Luce (1977) structured his comments around L. L. Thurstone's seminal work (1927a, b) on random variable models of human decision making. Luce's overview seems to us as relevant today as it was in 1977, so we use it here to clarify the constructs, assumptions, and interpretations of parameters in standard models of judgment and choice.

Because of the computational limitations of the early 20th century, Thurstone modeled only binary choices. His basic model equation was the following, which was defined at the individual level:

$$S_1 - S_2 = x_{12}(\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2)^{1/2}, \quad (1)$$

where  $S_j$  was the *psychological scale* value of stimulus  $j$ ,  $\sigma_j$  was the *discriminal dispersion* of stimulus  $j$ ,  $r$  was the correlation between the discriminational deviations of the two stimuli, and  $x_{12}$  was the standard normal deviation estimated from the observed proportion of choices that favored stimulus 1 over 2. Thurstone went on to identify the following key simplifying assumptions that made estimation possible for different types of data: (1) that the decision rule is to choose stimulus 1 over stimulus 2 whenever  $S_1 - S_2 > 0$ , (2) that differences between individuals in scale values for a given stimulus are normally distributed, (3) that  $r = 0$ , and (4) that  $\sigma_j = \sigma$  for all  $j$ .

Thurstone's model, of course, is the general binomial probit model that is still widely used today and is used in its multinomial form by Salisbury and Feinberg (2009). Importantly, Thurstone's vision of the process and his interpretation of the parameters are in some ways unchanged and in some ways rather different in current psychological models of human decision making (Luce 1977, McFadden 2001). It is still true that most researchers view decision making as based on noisy signals being processed in the brain that result in momentary *psychological impressions* of each choice alternative, and they find it useful to separate parameters that represent the amount of *internal noise*, those that represent individually stable *psychological representations* of choice alternatives,

and those that represent variation across individuals in internal noise or psychological representations (i.e., *heterogeneity*). What has changed since 1927 is that the decision rules themselves are modeled as random variables, and a single decision is modeled as the result of a number of micro-level stochastic processes that can operate sequentially or in parallel.

The most common form of the first change is found in variations of the theory of signal detection (Green and Swets 1966, Wixted 2007, Yonelinas and Parks 2007), in which the decision criterion is not assumed to be zero but is itself a random variable that is affected by internal noise (e.g., channel capacity; see Miller 1956, Wickens 2002) and stable psychological representations (e.g., incentives and other response biases; see Wixted 2007, Yonelinas and Parks 2007). For example, if choice options are considered sequentially, there may be a response bias toward choosing the first option (i.e., the decision criterion might be less than zero).

The most common form of the second change is found in various timing, counting, random walk, and neural net models (e.g., Busemeyer 1985; Busemeyer and Townsend 1993; Huang and Hutchinson 2008; Otter et al. 2008a, b; Townsend and Ashby 1983). Thus, a “Thurstonian” model is one that assumes that each decision can be represented by a single draw from a random distribution for each choice alternative and the decision rule is to choose the alternative with the largest value (i.e., a classic random utility model (RUM)). We note that early RUMs in econometrics interpreted the stochastic parameters as representing unobserved heterogeneity (e.g., McFadden 1974), whereas early psychometric models de-emphasized heterogeneity and interpreted stochastic parameters as internal noise (e.g., Luce 1959; Tversky 1972a, b).

From this perspective, the model of Salisbury and Feinberg (2008) is clearly a Thurstonian model. This is by no means a criticism! This level of modeling is a crucial link between current models of elementary decision making in mathematical psychology and current econometric models of market-level and macroeconomic data. We describe the more micro level of modeling because it informs our interpretation of Thurstonian parameters. In Salisbury and Feinberg’s (2009) model, only observed heterogeneity is modeled, so the stochastic parameters correspond to internal noise. Both  $\theta_j$  and  $\sigma_c$  are described as representing “consumers’ uncertainty,” which sometimes seems to mean beliefs that consumers have about their predictions of future experienced utility. It seems to us that beliefs about uncertainty, as they are commonly conceived in psychology and economics, are better captured as psychological representations than as internal noise. For example, most people are highly uncertain about the outcome of tossing a fair coin, but

there is very little internal noise affecting their belief that heads and tails are equally likely (e.g., the test-retest reliability is high for the subjective probability of the coin coming up heads). This perspective is consistent with structural modeling methods, which are typically very careful in separating beliefs about uncertainty and the belief updating process from variations in product quality, the internal noise affecting experienced utility, and heterogeneity (e.g., Erdem and Keane 1996, Erdem et al. 2004).

### Three Models of Judgment and Choice

To assess alternative interpretations of the temporal inflation parameter,  $\sigma_c$ , it is useful to introduce three variations of the basic Thurstonian model used by Salisbury and Feinberg (2009). The first model is specified as in Salisbury and Feinberg. Psychological impressions of utility,  $U_{ijt}^{(1)}$ , are represented as a combination of a stable representation of value,  $V_{ijt}^{(1)}$ , and internal noise,  $\theta_j \sigma_c^{(1)} \xi_{ijt}$ , in which temporal inflation is represented as a scale factor for the stochastic error term ( $\sigma_c^{(1)}$ ):

$$U_{ijt}^{(1)} = V_{ijt}^{(1)} + \theta_j \sigma_c^{(1)} \xi_{ijt}, \quad (2a)$$

and

$$V_{ijt}^{(1)} = \sum_{k=1}^K \beta_k X_{ijk}, \quad (2b)$$

where  $i$  indexes individuals,  $j$  indexes choice alternatives,  $t$  indexes time periods,  $k$  indexes the attributes (or other fixed effects) of each choice alternative (e.g.,  $RATE_{ij}$  and  $BRAND_j$  in the analyses reported by Salisbury and Feinberg 2009),  $c$  is *IMM* when  $t = 1$  and *FUT* when  $t > 1$  (see Salisbury and Feinberg for definitions of *IMM* and *FUT*), and  $\xi_{ijt}$  is a random variable representing stochastic error.

The second model treats temporal inflation as part of the psychological representation of the value of each choice alternative:

$$U_{ijt}^{(2)} = V_{ijt}^{(2)} / \sigma_c^{(2)} + \theta_j \xi_{ijt}, \quad (3a)$$

and

$$V_{ijt}^{(2)} = \sum_{k=1}^K \beta_k X_{ijk}. \quad (3b)$$

For the purposes of computing choice probabilities, the first and second models are equivalent because they both predict that the probability of choosing alternative A over alternative B is

$$P_{AB} = \Pr[\theta_B \xi_{iBt} - \theta_A \xi_{iAt} < (V_{iAt} - V_{iBt}) / \sigma_c] \quad (4)$$

(see Equation (2) in Salisbury and Feinberg 2009). Thus, for choice data, this indeterminacy must be resolved by assumption.

Salisbury and Feinberg’s (2009) interpretation of  $\sigma_c$  as internal noise is certainly plausible. However, note that the standard economic model of discounted utility (Samuelson 1937, see Frederick et al. 2002) defines the present value,  $PV$ , of a consumption experience that will occur  $\tau$  units of time in the future and provide utility value,  $V$ , as follows:

$$PV = V/(1 + \rho)^\tau, \quad (5)$$

where  $\rho$  represents the individual’s discount rate. It is easy to see that if we let  $\sigma_c = (1 + \rho)^\tau$ , then model 2 is implied without any effect of time on internal noise (defining  $\tau = 0$  when  $c = IMM$  and  $\tau = 1$  when  $c = FUT$ ).

Salisbury and Feinberg (2009) do not discuss economic models of present value; however, in their final discussion they say, “Such variance inflation factors are, of course, distinct in magnitude and concept from time-discounting parameters common in studies of intertemporal choice” (p. 15). We certainly agree that they are different in concept, both for the reasons just discussed and because many (but not all) intertemporal choice experiments involve decisions between alternatives that differ in temporal distance (e.g., \$75 now versus \$100 in one year). However, the estimated parameters from such experiments are not different in magnitude. Because the temporal distance in the reported experiment is one or two weeks, a rather large annual discount rate is implied by the estimated values of  $\sigma_c$ . Large (sometimes infinite) implied discount rates are not uncommon, especially when computed from short time delays (e.g., see the review by Frederick et al. 2002). Regardless of the implied magnitude, there is nothing in the reported analyses that rules out time discounting as an alternative explanation. 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 (i.e., value declines much more rapidly from the present to the near future than between more distant time points, in violation of the predictions of a constant discount rate; e.g., see Ainslie and Haslam 1992, Thaler 1981). We note, however, that models 1 and 2 both predict that  $\sigma_2 < \sigma_3$ .

Although models 1 and 2 cannot be differentiated using only choice data, they do make distinct predictions about judgment data, such as preference ratings, when it is reasonable to assume that the same evaluation process is used in both tasks and only the response component differs.<sup>1</sup> For example, assuming

that preference ratings,  $R_{ijt}$ , are a linear function of psychological impressions yields the following equations for models 1 and 2, respectively,

$$R_{ijt}^{(1)} = b_0 + b_1(V_{ijt}^{(1)} + \theta_j \sigma_c^{(1)} \xi_{ijt}), \quad (6a)$$

and

$$R_{ijt}^{(2)} = b_0 + b_1(V_{ijt}^{(2)} / \sigma_c^{(2)} + \theta_j \xi_{ijt}). \quad (6b)$$

For judgment data, model 1 predicts that estimates of error variance should increase with temporal distance, and model 2 predicts that they should remain constant. In the next section, we report an empirical test of these predictions.

The third model is fundamentally different from the first two because temporal inflation does not operate uniformly for all attributes. Instead, some attributes may be more heavily discounted than others, which is represented by  $\sigma_{ck}^{(3)}$  in the following model:

$$U_{ijt}^{(3)} = V_{ijt}^{(3)} + \theta_j \xi_{ijt}, \quad (7a)$$

and

$$V_{ijt}^{(3)} = \sum_{k=1}^K \beta_k X_{ijk} / \sigma_{ck}^{(3)}. \quad (7b)$$

Evidence favoring such *differential discounting* is a challenge for the first two models. Without differential discounting (i.e.,  $\sigma_{ck}^{(3)} = \sigma_c$  for all  $k$ ), model 3 reduces to model 2.<sup>2</sup> Note that Salisbury and Feinberg’s (2009) MNP-BF model allows for differential temporal effects for each brand. This model is a special case of model 3, in which  $\sigma$  is interpreted deterministically (like model 2) and each alternative is represented by a unique brand-specific discrete attribute (e.g., the unique aspects of Tversky’s elimination-by-aspects (EBA) model, 1972a, b).

In the next section, we report the results of experiments that test the hypothesis of differential discounting as an alternative to internal noise and the traditional concept of temporal discounting.

## Some Empirical Tests of the Interpretation of the Scale Parameter

It is beyond the scope of this commentary to fully test the ability of each of the three models described earlier to account for all known results or to conduct new experiments designed to differentiate between the models. However, to illustrate the types of tests that are possible and to provide a preliminary empirical assessment, we reanalyzed data from three papers

<sup>1</sup> This is certainly not universally true (e.g., Slovic and Lichtenstein 1983, Tversky et al. 1988); however, it is a reasonable assumption for some experimental paradigms, such as when both measures are collected simultaneously.

<sup>2</sup> Of course, attribute-specific error terms can be assumed in several different ways to create hybrid models (e.g.,  $V_{ijt}^{(4)} = \sum_{k=1}^K \beta_k (X_{ijk} + \xi_{ijkt}) / \sigma_{ck}^{(4)}$  or  $V_{ijt}^{(5)} = \sum_{k=1}^K \beta_k (X_{ijk} / \sigma_{ck}^{(5)} + \lambda_{ck}^{(5)} \xi_{ijkt})$ ). The exact specification would determine the net effects of temporal distance.

previously published by one of the authors of this paper (i.e., Malkoc et al. 2005, Zauberger and Lynch 2005, Zhao et al. 2007). Each of these papers reports simultaneous choice and relative preference ratings for two products (e.g., a 100-point scale in which 0 means “strongly favor A,” 50 means “indifferent between A and B,” and 100 means “strongly favor B”) that were made by two groups of subjects: one for whom the temporal distance of consumption was near and one for whom the temporal distance was far. In most cases, ratings favoring A or B were explicitly defined as choices, and subjects sometimes actually received their chosen product at the specified time. Such data seem to be as close to a simple stated choice task as a ratings task can be and, therefore, suitable to test the prediction of model 1, that internal noise increases with temporal distance. Assuming that heterogeneity does not increase with temporal distance, model 1 predicts that the variance in ratings should be greater for the temporally distant judgments than for the temporally near judgments.<sup>3</sup> In fact, the ratio of the standard deviations,  $SD_{FUT}/SD_{IMM}$ , is an estimate of  $\sigma_{FUT}^{(1)}$ . To the extent that immediate versus future is the dominant factor (and distance into the future is relatively unimportant as Salisbury and Feinberg found for their data), the estimated value of  $\sigma_{FUT}^{(1)}$  should be about two and, at a minimum, should be reliably greater than one. Models 2 and 3 predict there should be no effect of temporal distance and  $SD_{FUT}/SD_{IMM}$  should be approximately one.

Malkoc et al. (2005) report results from two experiments that share several key features with the Salisbury and Feinberg experiment. In their Experiment 1, participants were presented with a choice situation and asked to imagine choosing popcorn for a party in the near future (tomorrow night) or the distant future (six months from now). They were then provided with descriptions of two popcorn options on 12 attributes (adapted from Zhang and Markman 2001). The two options were pretested to be equal in overall attractiveness and favorability. However, they were also designed so that one product was more favorable on shared attributes (e.g., “large-sized kernels” versus “medium-sized kernels”) and the other was more favorable on unique attributes (“tastes a bit sweet” versus “not likely to burn”). Shared attributes provide a concrete basis for comparison, but unique attributes require the decision maker to integrate the information into a more abstract evaluation. Psychological theory predicts that because people process information for current decisions more

concretely, they will rely more heavily on the shared attributes for a temporally near judgment (e.g., Trope and Liberman 2003). Yet because they process information for future decisions more abstractly, they will rely on both shared and unique attributes for a temporally distant judgment. Thus, relative preference should conform to model 3. That is,  $\sigma_{IMM,k}^{(3)}$  will be larger for unique attributes than for shared attributes, but  $\sigma_{FUT,k}^{(3)}$  will be the same for all attributes, and there should be no effect of temporal distance on error variance.

The observed relative preference ratings supported this prediction. The mean number of points allocated to the brand favored by unique attributes was greater in the distant future ( $M = 43.6$ ,  $SD = 19.1$ ) than in the near future ( $M = 37.8$ ,  $SD = 18.6$ ),  $F(1, 175) = 4.33$ ,  $p < 0.05$ , and the standard deviations were approximately equal. This shift in preference without a shift in variance supports models 2 and 3. When examining the inferred choice data, we see a similar shift with the brand favored by unique attributes being chosen more often in the distant-future condition (30%) than in the near-future condition (21%). This difference in choice proportions is consistent with all three models but was not statistically significant;  $\chi^2(1, N = 151) = 1.76$ ,  $p = 0.18$ .

The second experiment from Malkoc et al. (2005) was similar in structure to Experiment 1 described above, but had participants rate and explicitly choose between two brands of potato chips, which they would actually receive, either in the near future (“at the end of this session”) or in the distant future (“at the end of the semester”). Again, the two brands were pretested to be equal in overall attractiveness and favorability. As before, the relative preference for the brand favored by unique attributes showed a significant effect of temporal distance,  $F(1, 149) = 4.58$ ,  $p < 0.05$ , indicating a greater preference for the brand favored by unique attributes in the distant future ( $M = 59.6$ ,  $SD = 17.4$ ) than in the near future ( $M = 53.4$ ,  $SD = 17.3$ ). It is noteworthy that, again, this shift in preference was not associated with changes in error variance, supporting models 2 and 3. The explicit binary choice data indicate that the brand favored by unique attributes was chosen more often in the distant-future condition (76%) than in the near-future condition (63%);  $\chi^2(1, N = 151) = 2.73$ ,  $p < 0.10$ . This result is not consistent with models 1 and 2 because they both predict that choice proportions should move closer to 0.5 as temporal distance increases. Differential weighting of attributes, however, predicts that the choice proportions for the brand favored by unique attributes should increase with temporal distance, regardless of its value when temporally near, consistent with model 3. This “main effect” shift with temporal distance (rather than a shift

<sup>3</sup> More specifically, model 1 predicts that the variance in ratings of relative preference for a given pair of alternatives should increase unless temporal distance also decreases heterogeneity and the magnitude of this decrease is greater than the increase predicted by model 1.

toward 0.5) is common in the intertemporal research literature (e.g., Liberman and Trope 1998, Trope and Liberman 2000, Soman 1998, Zauberman 2003) and is predicted by construal level theory (Trope and Liberman 2003), which has become increasingly influential in consumer decision research (Lynch and Zauberman 2007).

Of course, these results are mainly illustrative and experiments specifically designed to tease apart the various models (including hybrid combinations) will be needed to settle the issue. However, other research in psychology and behavioral economics has provided results consistent with the differential discounting of model 3 (e.g., Soman 1998; Trope and Liberman 2000, 2003; Zauberman 2003). We noted earlier that the ratio of standard deviations from relative preference tasks,  $SD_{FUT}/SD_{IMM}$ , is an approximate estimate of  $\sigma_{FUT}^{(1)}$ . Table 1 provides estimates based on the standard deviations reported in Malkoc et al. (2005), Zauberman and Lynch (2005), and Zhao et al. (2007), which all measured relative preference in temporally near and far conditions. We have also added data from two other published papers, one (Trope and

Liberman 2000) that reports only ratings and standard deviations and one (Zauberman 2003) that includes only choice proportions. Overall, these estimates of  $\sigma_{FUT}^{(1)}$  are all very close to one with no systematic tendency to be greater than one, let alone close to two. Also, choice proportions do not always move toward 0.5 for future consumption; sometimes they become more extreme and sometimes they reverse (i.e., cross the 0.5 boundary). Taken as a whole, these results are inconsistent with Salisbury and Feinberg's account of the effect of temporal distance.<sup>4</sup>

In sum, although these results are obviously not comprehensive and are not intended to serve as a formal meta-analysis, we believe that they do reflect the overall pattern observed in the literature on intertemporal decisions. We think that there are two key lessons that these data illustrate: the first is that we do not observe increases in standard deviations for preference ratings in the future compared to the present, and the second is that the shift in choice proportions does not necessarily regress to 50/50, but rather that the typical results reflect a main effect of differential discounting, increased future choice proportions for alternatives with more concrete costs and more abstract benefits compared to other alternatives (e.g., Trope and Liberman 2003).

**Table 1** Empirical Tests of Salisbury and Feinberg's (2009) Model Found in Previously Published Research

Source	Choice <sub>IMM</sub>	Choice <sub>FUT</sub>	SD <sub>IMM</sub>	SD <sub>FUT</sub>	$\sigma_{FUT}^{(1)}$
MZU, Exp. 1	0.21 <sup>-</sup>	0.30 <sup>+</sup>	18.6	19.1	1.03
MZU, Exp. 2*	0.63 <sup>-</sup>	0.76 <sup>+</sup>	17.3	17.4	1.01
ZL, Exp. 5, Cond = T	0.39 <sup>-</sup>	0.56 <sup>+</sup>	2.4	2.4	0.99
ZL, Exp. 5, Cond = M*	0.70 <sup>-</sup>	0.63 <sup>+</sup>	2.5	2.4	0.96
ZHZ; Exp. 1, Cond = C	0.35 <sup>-</sup>	0.41 <sup>+</sup>	2.4	2.0	0.83
ZHZ; Exp. 1, Cond = O*	0.52 <sup>+</sup>	0.48 <sup>+</sup>	2.2	2.6	1.21
ZHZ; Exp. 1, Cond = P*	0.18 <sup>-</sup>	0.23 <sup>-</sup>	2.3	2.3	1.00
ZHZ; Exp. 2, Cond = C	0.28 <sup>-</sup>	0.58 <sup>+</sup>	2.6	3.0	1.15
ZHZ; Exp. 2, Cond = O*	0.51 <sup>+</sup>	0.56 <sup>+</sup>	2.8	2.6	0.93
ZHZ; Exp. 2, Cond = P*	0.22 <sup>-</sup>	0.37 <sup>-</sup>	2.7	2.6	0.97
TL; Exp. 2, Cond = HLP	—	—	1.6	1.7	1.07
TL; Exp. 2, Cond = LLP	—	—	1.9	2.1	1.10
TL; Exp. 3, Cond = HLP	—	—	2.3	2.2	0.98
TL; Exp. 3, Cond = LLP	—	—	2.0	1.8	0.93
TL; Exp. 4, Cond = HLP	—	—	2.1	1.2	0.60
TL; Exp. 4, Cond = LLP	—	—	2.5	2.6	1.06
Z; Exp. 3, Cond = LSU*	0.71 <sup>-</sup>	0.83 <sup>+</sup>	—	—	—
Z; Exp. 3, Cond = HSU	0.29 <sup>-</sup>	0.58 <sup>+</sup>	—	—	—

*Notes.* References used are as follows: MZU refers to Malkoc et al. (2005); TL refers to Trope and Liberman (2000); Z refers to Zauberman (2003); ZL refers to Zauberman and Lynch (2005); ZHZ refers to Zhao et al. (2007). "Exp" refers to specific experiments in each article, and "Cond" refers to a specific condition in a given experiment. T = time; M = money; C = control; O = outcome simulation; P = process simulation; HLP = high construal level positive; LLP = low construal level positive; LSU = low setup; HSU = high setup. The predictions of the differential discounting model are indicated by pluses (+, larger proportion) and minuses (-, smaller proportion); note that ZHZ predicted a specific pattern across conditions that sometimes included a prediction of no effect for temporal distance. An asterisk (\*) indicates experimental conditions for which the authors' predictions for choice proportions (i.e., the predictions of differential discounting) differed from those of temporal inflation.

## Conclusions and Future Research

Although we may disagree with Salisbury and Feinberg (2009) about which model (1, 2, or 3) best describes the process that drives present versus future preferences, we have no quarrel with the larger thrust of their paper: the idea that the effects of experimental variables on unobserved variance can distort estimated parameters in choice analysis. This is a caution that has been repeatedly voiced over the years by a number of others in the field, most notably by Louviere and his colleagues (e.g., Louviere 2001, Louviere and Meyer 2007), but to date it has gone largely unheeded. We hope that the appearance of this paper will serve the positive role of bringing greater attention to this issue and spark a constructive dialogue

<sup>4</sup> As was discussed earlier, Salisbury and Feinberg's (2009) most general model, MNP-BF, can capture differential discounting in the special case where each brand is defined by a unique feature (e.g., a special ingredient) rather than an array of values on a battery of generic attributes (e.g., degree of sweetness, cost, etc.). If there had been strong effects of differential discounting across features in their data, then the MNP-BF model should have fit better than the MNP-BT model, but it did not. This is to be expected because their alternatives (snacks) did not differ on features that have been found to have different discount rates in past research, such as if some snacks were seen as more desirable versus more convenient than others (e.g., Trope and Liberman 2003). As discussed earlier, the main problem for their data, considered in isolation, is the inability to distinguish between models 1 and 2.

among scholars about its implications for how both behavioral and quantitative methods in choice analysis should be used and improved going forward.

However, we also worry that the positive message of this paper might be overlooked by some readers, with the work being misread as advocating for the use of statistical models of complex error structures as “default” theories of choice that must be rejected by behavioral theorists when conducting research. That is, before a behavioral researcher can claim support for a behavioral hypothesis about a judgment process he or she must first demonstrate that the data could not be equally well described by a random utility model that recognizes a generalized (heteroskedastic) error structure. In our view such a stance would misread the spirit of the arguments made by Louviere (2001) and Hutchinson et al. (2000). In our view both those prior efforts and the current paper raise the bar for *both* research camps—it challenges behavioral researchers to more deeply consider how unobserved variance might affect the interpretation of experimental results, but it also challenges quantitative researchers *not* to view statistical accounts of error structures as being sufficient explanations for the process of choice.

To elaborate, taken at face value, Salisbury and Feinberg (2009) impart a not-so-subtle slap on the wrist of behavioral researchers by arguing that the previous experimental work on intertemporal choice is flawed by its failure to explicitly consider the possibility that choice precision may vary over conditions. This critique is, of course, somewhat paradoxical since it was behavioral researchers who first pointed out the contextual nature of preferences and choices. Hence, presumably few in this group would contest a suggestion that the unobserved components of utility will also likely vary by context. Yet while behavioral researchers might anticipate this confounding in the abstract, it is a legitimate criticism that its key statistical implication has been overlooked: the fact that it potentially renders tests of hypotheses about process based on comparisons of aggregate means ambiguous. We should emphasize, however, that this critique does not apply to *all* behavioral findings: it uniquely applies to tests of behavioral theories where the core prediction is the deflation or amplification of a main effect across experimental treatments—an effect that could be artificially produced by unmeasured changes in the underlying variance.

We argue, however, that a finding of nonconstant error variance poses no less of a challenge for quantitative modelers, although of different form. The problem is this: One might read Salisbury and Feinberg’s (2009) paper as suggesting that a suitable fix for a finding of temporal variation in errors is simply to estimate a model that relaxes the assumption that errors are time

invariant. Note that although this indeed “fixes” the problem in a purely statistical sense, it does not remedy it in a behavioral sense. Specifically, rather than pushing analysts to try to explain *why* the variance of the unobserved component of utility varies over time, it lets them off the hook by suggesting that it is sufficient simply to describe the statistical properties of these errors. In our view the key challenge to modelers when faced with findings of temporal (or contextual) invariance is to *avoid* the temptation to accept such statistical descriptions as explanations. A finding of heteroskedasticity should be seen as evidence that the underlying assumptions that are being made about the process that drives choice are almost certainly wrong, and a search should begin for alternative, theoretically grounded models of both psychological representations and internal noise and their interactions.

In this paper we argued that the mere fact that a certain set of intertemporal choice data can be well fit by a time-invariant utility function, and a time-varying error term does *not* imply that this account explains the underlying process that is driving choices. As discussed earlier, the same results could arise from temporally discounted utilities and stationary precision. Although “uncertainty about future expected utility” could indeed be the proper explanation for the process that underlies the observed choices, its validity can be assessed only after it is rigorously tested against other possible process explanations that could leave a similar data signature.

However, we recognize that such a goal of developing better theoretically grounded models may be easier to state than implement. An initial technical challenge that arises when testing unobserved variance explanations for choice data against alternative accounts is that empirical separation may often be difficult. For example, in their 2009 paper, Salisbury and Feinberg attempt to rule out variety seeking as an explanation of temporal inflation by showing that a single parameter representing the net effect of inertia and variety seeking was not significantly different from zero and did not affect the estimated value of  $\sigma_{FUT}^{(1)}$ . These results are indeed supportive of their claim; however, single-parameter representations of this type are known to confound multiple sources of state dependence and are influenced by heterogeneity (e.g., see Seetharaman 2004). These sources might overwhelm variety-seeking tendencies, if they were present. Future attempts to decouple the sources of state dependence may require experiments or statistical tests that are specifically designed for this purpose, such as those reported by Read and Lowenstein (1995). In a three-choice paradigm very similar to that of Salisbury and Feinberg, they examined data from subjects who chose exactly two different snacks and observed the relative frequency of the ABA sequence

compared to the AAB and ABB sequences as an indicator of variety seeking.<sup>5</sup> In sequential conditions, the percentage choosing ABA ranged from 21% to 35% and in the simultaneous conditions (where variety seeking was hypothesized to be influential) it ranged from 55% to 75%. This supports variety seeking and cannot be explained by variance inflation because  $\sigma_2^{(1)}$  should be less than or equal to  $\sigma_3^{(1)}$ , and this implies that AAB should be as or more likely than ABA in the simultaneous conditions (given the plausible assumption that  $V_A > V_B$ ). Salisbury and Feinberg estimated but did not report models in which the constraint that  $\sigma_{FUT}^{(1)} = \sigma_2^{(1)} = \sigma_3^{(1)}$  was relaxed. It would be useful to know what those estimates were because variety seeking suggests that  $\sigma_2^{(1)}$  can be greater than  $\sigma_3^{(1)}$ , contrary to the hypothesis of temporal inflation. Of course, the more direct test would be to see whether or not their experiment replicated the results of Read and Lowenstein for the ABA sequence.

There is, however, an even deeper, and perhaps more troubling, challenge to the study of choice: the growing separation that exists between quantitative and behavioral researchers who share a common interest in the study of decision making. Although, in principle, quantitative researchers could benefit from the insights into the psychology of choice that could be provided by behavioral researchers, such collaboration is surprisingly rare. One contributing factor is that in recent years the field of formal choice modeling has become dominated by the study of probabilistic error structures—a shift that has made its content and findings remote from the interests of behavioral researchers. Whereas behavioral researchers, for their part, could benefit from advances in modeling and analysis methods developed by their quantitative counterparts (as Salisbury and Feinberg 2009 illustrate), few are given the training needed to absorb or implement these advances. In addition, behavioral research has been increasingly dominated in recent years by the study of problems emerging from social psychology, a field that is both conceptually and methodologically remote from the fields of formal choice analysis and mathematical psychology. However, it is clear that if we are to make real progress in the understanding of choice behavior it will not happen with quantitative and behavioral research camps going it alone. Behavioral researchers *should* be paying closer attention to the statistical issues Salisbury and Feinberg raised in their paper, and quantitative researchers *should* be focusing on building models that provide a more complete and realistic

account of the actual behavioral process that underlies choice. Our hope is that the commentaries triggered by the Salisbury and Feinberg paper act to encourage greater dialogue between behavioral and quantitative research camps, and reverse the trend of separation that has existed in recent years.

## References

- Ainslie, G., N. Haslam. 1992. Hyperbolic discounting. G. Lowenstein, J. Elster, eds. *Choice Over Time*. Russell Sage Foundation, New York, 57–92.
- Brenner, L. A. 2000. Should observed overconfidence be dismissed as a statistical artifact? Critique of Erev, Wallsten, and Budescu (1994). *Psych. Rev.* **107**(4) 943–946.
- Busemeyer, J. R. 1985. Decision making under uncertainty: A comparison of simple scalability, fixed-sample and sequential-sampling models. *J. Experiment. Psych.: Learn., Memory, Cognition* **11**(3) 538–564.
- Busemeyer, J. R., J. T. Townsend. 1993. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psych. Rev.* **100**(3) 432–459.
- Erdem, T., M. P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15**(1) 1–20.
- Erdem, T., Y. Zhao, A. Valenzuela. 2004. Performance of store brands: A cross-country analysis of consumer store-brand preferences, perceptions, and risk. *J. Marketing Res.* **41**(1) 86–100.
- Erev, I., T. S. Wallsten, D. V. Budescu. 1994. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psych. Rev.* **101**(3) 519–527.
- Frederick, S., G. F. Lowenstein, T. O'Donoghue. 2002. Time discounting and time preference: A critical review. *J. Econom. Literature* **40**(2) 351–401.
- Green, D. M., J. A. Swets. 1966. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York.
- Huang, Y., J. W. Hutchinson. 2008. Counting every thought: Implicit measures of cognitive responses to advertising. *J. Consumer Res.* **35**(1) 98–118.
- Hunter, J. E., F. L. Schmidt. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage Publications, Thousand Oaks, CA.
- Hutchinson, J. W., W. A. Kamakura, J. G. Lynch, Jr. 2000. Unobserved heterogeneity as an alternative explanation for “reversal” effects in behavioral research. *J. Consumer Res.* **27**(3) 324–344.
- Liberman, N., Y. Trope. 1998. The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *J. Personality Soc. Psych.* **75**(1) 5–18.
- Lichtenstein, S., B. Fischhoff, L. D. Phillips. 1982. Calibration of probabilities: The state of the art to 1980. D. Kahneman, P. Slovic, A. Tversky, eds. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 306–334.
- Louviere, J. J. 2001. What if consumer choice experiments impact variances as well as means? Response variability as a behavioral phenomenon. *J. Consumer Res.* **28**(3) 506–511.
- 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. *Rev. Marketing Res.* **4** 3–32.
- Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons, New York.
- Luce, R. D. 1977. Thorstone's discriminational processes fifty years later. *Psychometrika* **42**(4) 461–489.

<sup>5</sup> Only three sequences are possible when three choices select exactly two items. Labeling the first choice “A,” these are AAB, ABA, and ABB. Note that this analysis controls for heterogeneity to a considerable degree because the first choice varies across individuals.



- Lynch, J. G., G. Zauberman. 2007. Construing consumer decision making. *J. Consumer Psych.* **17**(2) 107–112.
- Malkoc, S., G. Zauberman, C. Ulu. 2005. Consuming now or later? The interactive effect of timing and attribute alignability. *Psych. Sci.* **16**(5) 411–417.
- McFadden, D. F. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed. *Frontiers of Econometrics*. Academic Press, New York, 104–142.
- McFadden, D. F. 2001. Economic choices. *Amer. Econom. Rev.* **91**(3) 351–378.
- Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psych. Rev.* **63**(March) 81–97.
- Otter, T., G. M. Allenby, T. van Zandt. 2008a. An integrated model of discrete choice and response time. *J. Marketing Res.* **45**(5) 593–607.
- Otter, T., J. Johnson, J. Rieskamp, G. M. Allenby, J. D. Brazell, A. Diederich, J. W. Hutchinson, S. MacEachern, S. Ruan, J. Townsend. 2008b. Sequential sampling models of choice: Some recent advances. *Marketing Lett.* **19**(3–4) 255–267.
- Read, D., G. Lowenstein. 1995. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *J. Experiment. Psych.-Appl.* **1**(1) 34–49.
- Salisbury, L. C., F. M. Feinberg. 2008. Future preference uncertainty and diversification: The role of temporal stochastic inflation. *J. Consumer Res.* **35**(2) 349–359.
- Salisbury, L. C., F. M. Feinberg. 2009. Alleviating the constant stochastic variance assumption in decision research: Theory, measurement, and experimental test. *Marketing Sci.*, ePub ahead of print November 3, <http://mktsci.journal.informs.org/cgi/content/abstract/mksc.1080.0464v1>.
- Samuelson, P. A. 1937. A note on measurement of utility. *Rev. Econom. Stud.* **4**(2) 155–161.
- Seetharaman, P. B. 2004. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Sci.* **23**(2) 263–271.
- Slovic, P., S. Lichtenstein. 1983. Preference reversals: A broader perspective. *Amer. Econom. Rev.* **73**(4) 596–605.
- Soman, D. 1998. The illusion of delayed incentives: Evaluating future effort-money transactions. *J. Marketing Res.* **35**(4) 427–437.
- Thaler, R. 1981. Some empirical evidence of dynamic inconsistency. *Econom. Lett.* **8**(3) 201–207.
- Thurstone, L. L. 1927a. A law of comparative judgments. *Psych. Rev.* **34** 273–286.
- Thurstone, L. L. 1927b. Psychophysical analysis. *Amer. J. Psych.* **38** 368–389.
- Townsend, J. T., F. G. Ashby. 1983. *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge University Press, Cambridge, UK.
- Trope, Y., N. Liberman. 2000. Temporal construal and time-dependent changes in preference. *J. Personality Soc. Psych.* **79**(6) 876–889.
- Trope, Y., N. Liberman. 2003. Temporal construal. *Psych. Rev.* **110**(3) 403–421.
- Tversky, A. 1972a. Choice by elimination. *J. Math. Psych.* **9**(November) 341–367.
- Tversky, A. 1972b. Elimination by aspects: A theory of choice. *Psych. Rev.* **79**(4) 281–299.
- Tversky, A., S. Sattath, P. Slovic. 1988. Contingent weighting in judgment and choice. *Psych. Rev.* **95**(3) 371–384.
- Wallsten, T. S., I. Erev, D. V. Budescu. 2000. The importance of theory: Response to Brenner (2000). *Psych. Rev.* **107**(4) 947–949.
- Wickens, T. D. 2002. *Elementary Signal Detection Theory*. Oxford University Press, New York.
- Wixted, J. T. 2007. Dual-process theory and signal-detection theory of recognition memory. *Psych. Rev.* **114**(1) 152–176.
- Yonelinas, A. P., C. M. Parks. 2007. Receiver operating characteristics (ROCs) in recognition memory: A review. *Psych. Bull.* **133**(5) 800–832.
- Zauberman, G. 2003. The intertemporal dynamics of consumer lock-in. *J. Consumer Res.* **30**(3) 405–419.
- Zauberman, G., J. G. Lynch Jr. 2005. Resource slack and propensity to discount delayed investments of time versus money. *J. Experiment. Psych.: General* **134**(1) 23–37.
- Zhang, S., A. B. Markman. 2001. Processing product unique features: Alignability and involvement in preference construction. *J. Consumer Psych.* **11**(1) 13–27.
- Zhao, M., S. Hoefler, 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.