Comments

Information Overload and the Nonrobustness of Linear Models: A Comment on Keller and Staelin

ROBERT J. MEYER
ERIC J. JOHNSON*

In a recent article, Keller and Staelin (1987) reexamined one of the more controversial questions in consumer research literature: Is it possible to provide consumers with too much information when making choices? Their work contained the valuable and insightful conjecture that increasing quantity of information available to consumers increases the overall quality of information as well, when quality is defined as the cumulative importance of each bit of information. Because prior research had not explicitly controlled these two effects, degrading effects of increasing information quantity may have been muted by unintended, but commensurate, increases in information quality. Their research objective was to decompose these two sources of variance experimentally.

Keller and Staelin found that when information quality was held constant, increases in information had a strong negative effect on decision accuracy. When information quantity was held constant, increasing information quality had a positive effect, at least to a point. Keller and Staelin (1987, p. 212) interpreted their results as evidence that consumers can be overloaded with information.

> The results of this study imply that consumers are not able to shield themselves from being overloaded when "too much" information is made readily available to the decision maker; too much being related to the quantity and average quality of the available information.

Does the Keller and Staelin investigation close the book on the information overload debate? In our view, refinement of the relationship between information quality and quantity is valuable, but a strong argument can be built that their data do not offer an unambiguous demonstration of information overload. Specifically, it is possible to show that a similar set of overload results could have been obtained within their methodology even if consumers had an unlimited ability to process information.

Our critique centers on Keller and Staelin's approach to identifying consumer error in a given choice, specifically, their use of inferred choice errors. An inferred error arises when an analyst makes an assumption about how decision makers should combine attribute information in a given setting, usually a weighted linear composite of attributes. An error is any choice that does not agree with this prescription. This contrasts with consistency-based measures, which define errors more narrowly as only those choices that violate a primitive axiom of rationality (e.g., transitivity or the selection of dominated options). A central tenet of our discussion is that a fundamental indeterminacy influences any attempt to assess the optimality of choices based on inferred errors. These problems arise from inherent measurement error, which will often increase as a function of information quantity, and the natural instability and lability of preferences, which clouds distinctions between right and wrong choices.

*Robert Meyer is Associate Professor of Marketing at the Anderson Graduate School of Management, University of California, Los Angeles, CA 90024. Eric Johnson is Associate Professor of Marketing at the Wharton School, University of Pennsylvania, Philadelphia, PA 19104. The authors would like to acknowledge the support of a contract from the Office of Naval Research Engineering Psychology Program. They also express their gratitude to Rick Staelin (who offered them their first tenure-track faculty positions in marketing) and Kevin Keller (a long-time jogging partner) for their encouragement and comments.

INFORMATION OVERLOAD AND INFERRED CHOICE ERRORS

Keller and Staelin used what would seem to be a straightforward procedure to investigate how information quantity and quality affect decision accuracy. Subjects were asked to provide ratings of the relative importance of a number of potential job attributes and the attractiveness of differing potential levels of these attributes. These were taken as measures of the normative importance and value of differing attributes against which later actual choices would be compared. Two months later, a subset of the original subject pool was presented with sets of job descriptions constructed from these ratings. Central to this phase was a manipulation of information quantity, or the number of attributes by which a job was described, and information quality, defined as the sum of the stated importance of these attributes.

Keller and Staelin's interest was the extent to which the subjects made correct choices across varying information quantity and quality conditions. A correct choice was defined as the job that maximized the linear-weighted criterion:

$$u_i = \sum_{k=1}^{n} w_k e_{ki}, \tag{1}$$

where w_k was the importance of attribute k (on a 1–100 scale) as measured by the first stage questionnaire, and e_{ki} was a measure of the evaluation of option i on attribute k (on a 1–100 scale), also measured in the first stage.

Keller and Staelin recognized that their definition of normative choice might be fallible. Choices that departed from normative prediction may have reflected suboptimality on the part of subjects or error in Keller and Staelin's definition of the normative overall utility of each option, u_i. Such misspecification might have accrued to error in their measures of the relative importance of differing attributes among subjects (w_k), the attractiveness of the values of options on these attributes (e_{ki}), or a failure of their assumption that a linear-additive rule was the preferred integration strategy among subjects. The potential for error, however, was not considered an obstacle because their interest was in explaining variance in performance rates rather than absolute prediction levels. They assumed that measurement error was independent of the predicted value of u_i, and concluded that the observed effects of information quantity and quality reflected true variation in processing abilities, not measurement error.

We question this assumption, because Keller and Staelin's design posits a theoretical relationship between the likelihood of prediction errors and information quantity and quality that muddles the interpretability of their findings. Specifically, in many cases, one would expect inferred choice accuracy to decrease with increases in information quantity and increase with total quality, even if subjects had unlimited processing abilities.

The demonstration of this result is straightforward. Recall that in any choice experiment the likelihood that an analyst will make an error in predicting which option a subject will choose from a set will be affected by two potentially separate factors—the difficulty of the task, or the variance that exists between options, and the difficulty of measurement within the task, or the variance within options. These sources of variance can be problematic for information overload experiments like Keller and Staelin's because they will generally covary with the amount of information used to describe options and the average quality of this information. Specifically, increasing levels of information quantity will tend to be associated with more difficult choices and less precise measures of subjects' utilities, while increasing levels of total quality will tend to be associated with easier choices and a constant precision of measurement. Because variations in task ease and measurement error affect the analyst's ability to forecast which options would be most preferred by subjects, it is difficult to differentiate those choice errors that stem from suboptimality on the part of subjects from those that accrue to suboptimal predictions on the part of the analyst.

Changes in information quantity and quality are likely to affect ease of choice tasks faced by subjects, a circumstance that Keller and Staelin did attempt to control in their work. However, variations in quantity and quality may have induced variations in measurement error within tasks, a factor they did not (and likely could not) control.

Quality, Quantity, and Task Ease

Assume that a choice experiment presents subjects with a number of randomly generated pairs of job descriptions. In addition, assume that the overall utilities of each of these options have been measured by the analyst with an error that is independent of the given levels of information quantity and quality (Keller and Staelin's starting assumption). Our interest is in exploring how controlled changes in information quantity and quality affect the ease of the choice tasks faced by subjects and the likelihood that the analyst will make an error in identifying the subject's optimal choice.

Let δ_{ij} be a measure of the ease of the choice between options i and j, formally defined as the difference in their overall measured utility values. In particular, following Equation 1,

$$\delta_{ij} = \sum_{k=1}^{n} w_k(e_{ki} - e_{kj}), \tag{2}$$

where, w_k is again the analyst's estimate of the relative importance of the kth attribute to the subject (a constant for a given subject within any quantity/quality condition), and $(e_{ki} - e_{kj})$ is the difference in the evaluations of this attribute across stimuli in a condition (a random variable). Given a constant measurement error, it should be clear that the smaller the absolute value of δ_{ij}, the less certain the analyst can be regarding the normative best option for the subject.

Consider first the behavior of δ_{ij} as information quantity increases with total information quality held constant. If we assume that weights are equal across attributes within any quantity level, the attribute weight in each condition will be given by w_k = (total quality)/n. Thus, as n increases, the average weight decreases. Substituting this definition of w_k, Equation 2 may be rewritten as follows:

$$\delta_{ij} = \sum_{k=1}^{n} \frac{\mathrm{TQ}}{n}(e_{ki} - e_{kj}). \quad (3)$$

The likely effect of changes in n (holding total quality constant) on task difficulty can be inferred by computing the theoretical means and variances of δ_{ij}. If i or j remains equally attractive across trials (i.e., the task is unbiased), and attribute evaluations have equal variances, it follows that

$$E(\delta_{ij}) = 0 \quad (4)$$

and

$$\mathrm{VAR}(\delta_{ij}) = \frac{\mathrm{TQ}^2}{n}(2\mathrm{VAR}(e)), \quad (5)$$

where $\mathrm{VAR}(e^*)$ is the variance in the evaluations of an attribute across trials.

The moments of Equations 4 and 5 show that when total quality is held constant, increases in quantity decreases the absolute difference in overall values between any two options. This converges to zero, implying a systematic increase in the difficulty of the task. Given independent measurement error in overall utilities, it appears that increases in information quantity, holding total quality constant, will produce decreases in the accuracy of the analyst's predictions. When total quality is varied and quantity is held constant, the reverse is the case, and there is an increase in the relative number of easily discriminable choices as quality increases. The basic rationale is that differences that exist between options on their attributes become amplified when multiplied by increasingly large scaling factors, in this case importance weights. Formally, if s is a scaling constant reflecting differences in the average importance of attributes across conditions, the variance in utility differences given in Equation 5 can be rewritten:

$$\mathrm{VAR}(\delta_{ij}) = \frac{s\mathrm{TQ}^2}{n}(2\mathrm{VAR}(e)). \quad (6)$$

As s increases (i.e., attributes become, on average, more important), the variability in utility differences increases, implying a greater relative frequency of easy choice problems from the analyst's perspective. We would thus expect an increase in decision accuracy as the average importance of the attributes presented to subjects grows.

Quantity, Quality, and Measurement Error

To control for any confounding effect of variations in task difficulty, the Keller and Staelin analysis included a task ease covariate, which was the difference in the predicted overall utility values between the prescribed best and second best options in each choice set. Although likely fallible in allowing a residual influence of variations in task difficulty, this control would seem sufficient nevertheless to rule out variations in task ease as the sole explanation for their results. Variation in task ease, however, is only one potential determinant of prediction errors. Also relevant is the variance within options, or the precision of the estimates of overall utility. We now consider how variations in information quantity and quality may have affected the precision of estimates of overall utility, holding the ease of the task constant.

We first consider the case where these measurement errors are independent and identically distributed across attribute weights and evaluations. We then consider the case where there are dependencies in errors, such as might arise if subjects exerted more or less care when judging attributes of varying importance.

Let ϵ_i be the analyst's error in the measurement of the true overall utility associated with option i ($\epsilon_i = u_i - u_i^*$), let ϵ_w be the error in measuring the true weight associated with any attribute ($\epsilon_w = w_k - w_k^* \; \forall k$), and ϵ_e be the error in measuring the true evaluation of an option on an attribute ($\epsilon_e = e_{ki} - e_{ki}^* \; \forall k, i$). ϵ_i, ϵ_w, and ϵ_e are assumed to be independent random variables with means 0 and variances $\mathrm{VAR}(\epsilon_i)$, $\mathrm{VAR}(\epsilon_w)$, and $\mathrm{VAR}(\epsilon_e)$, respectively. The overall measurement error for option i, ϵ_i, can be expressed in terms of ϵ_w and ϵ_e as follows:

$$\begin{aligned}\epsilon_i &= u_i - u_i^* \\ &= \sum_{k=1}^{n} (w_k^* + \epsilon_w)(e_{ki}^* + \epsilon_e) - \sum_{k=1}^{n} w_k e_{ki} \\ &= \sum_{k=1}^{n} (w_k^*\epsilon_e + e_{ki}^*\epsilon_w + \epsilon_w\epsilon_e)\end{aligned}$$

with a corresponding variance

$$\mathrm{VAR}(\epsilon_i) = \sum_{k=1}^{n} (w_k^*\mathrm{VAR}(\epsilon_e) + e_{ki}^*\mathrm{VAR}(\epsilon_w) + \mathrm{VAR}(\epsilon_w\epsilon_e)). \quad (7)$$

According to Equation 7, experimental increases in quantity will lead to less precise estimates of overall utility, holding total quality constant. In Keller and Staelin's design, increases in quantity were manipulated by increases in n, which increases overall error variance even when the average weight w_k decreases with n. Equation 7 also implies that the effect of increasing information quality could be affected by measurement error in the attribute evaluations, ϵ_{ki}. If, as Keller and Staelin assumed, the attribute values are measured accurately, ($\epsilon_e = 0$), changes in total quality should have no effect on the precision of estimates of overall utility when quantity is held constant. If there is measurement error in the attribute evaluations, however, increases in quality will decrease precision.

This analysis assumes that measurement errors are independent of the given attribute score or weight. In Keller and Staelin's study, however, this may not have been the case because subjects may have taken more care in evaluating more important attributes, causing measurement error to decrease with increases in attribute weight. Also, Keller and Staelin report some evidence that errors may have been somewhat larger for more important attributes (1987, p. 207, footnote).

Imposing linear covariation in measurement error and weight judgments directly affects the likelihood of errors: If errors are smaller for more important attributes, the effects of information quantity implied by Equation 7 would be amplified. Because increases in quantity are associated with decreases in the average attribute importance, the overall measurement error would include a larger sum of errors as well as a larger average size of errors. No error in the measurement of attribute scores (ϵ_e) would also have a positive effect on information quality because as average quality increases, holding quantity constant, predictions of utility would be based on an increasingly precise set of attribute weight measures. In contrast, if measurement errors are larger for more important attributes, as Keller and Staelin argued, we would expect a mollification of the effect of quantity due to cancellation; increases in total measurement error accruing to a larger sum of errors would tend to be offset by increased precision of each weight measure.

A MONTE CARLO EXPERIMENT

The previous analysis suggests that there are theoretical reasons for suspecting that Keller and Staelin's conclusions about the effects of information quantity and quality may have an artifactual basis. We have to temper this assertion, however, for two reasons. First, as noted, Keller and Staelin's analysis did attempt to control for at least one of the natural sources of variation in prediction error, task ease. Second, there may have been a positive correlation between the amount of measurement error associated with an attribute and its judged importance. Both factors would tend to mollify artifactual explanations for their findings.

To explore the effect of these factors on predicted choice errors, we undertook the following numerical simulation. Following Keller and Staelin's procedure, we conducted three replications of a Monte Carlo experiment in which 100 subjects made 100 pairwise choices that varied by number of attributes and total importance weight associated with those attributes. Formally, we assumed that the true utility associated with an option in a choice set was given by the linear composite represented in Equation 1. We assumed that there was no error in our measurement of attribute scale values $e_{kj(i)}$ (a conservative assumption vis-à-vis Keller and Staelin) but that there was error in the weight measures. This error was modeled as an independent, identically distributed, uniform random variate with mean 0 and range r.

Our interest was the ability of models given a set of fallible weights to predict the choices made by a set of unobserved true weights. We examined this accuracy varying information quantity, quality, weight noise, and the degree of covariation between weight noise and the size of a given weight measure. Within the independent weight noise condition, we examined information quantity (4, 8, 10, or 12 attributes), total information quality (360, 510, 630, or 800 cumulative importance points), and noise level (w $\pm$ 10, $\pm$30, or $\pm$50 importance points). In the correlated weight noise condition, we examined the same levels of quantity and quality. The pattern of covariation was one in which the error size increased linearly from $\pm$10 importance points for weights of 0 to 10 to $\pm$50 importance points for weights of 100 or more. Because of this truncation, the correlation between weight and error was not a perfect one, but was 0.47 across all quantity and quality levels.

These levels of quantity and quality mirror those used by Keller and Staelin,[1] and the levels of independent error were designed to reflect a plausible range for their task. The largest error level (50 percent) simulated a case with significant error in measuring both attribute weights and scale values in the task, and was perhaps larger than the level that existed in Keller and

[1]In the Keller and Staelin experiment, weights were strictly bounded between 1 and 100. In our simulation, weights greater than 100 occurred in those conditions (not feasible in the Keller and Staelin design) where total quality was at a high level (e.g., 800 cumulative points) and quantity at a small level (e.g., four attributes). Weights less than one were possible in the simulation in instances where the mean weight value was less than the uniform error range. In earlier versions of the simulation, we imposed a constraint preventing negative weights, with results similar to those reported here.

TABLE

REGRESSION COEFFICIENT ESTIMATES, MONTE CARLO SIMULATION OF KELLER AND STAELIN'S EXPERIMENT

Variables	Independent errors	Errors correlated with weights
Log(N)	$-.5138 \times 10^{-1}$ ($p < 0.001$)	$-.5148 \times 10^{-1}$ ($p < 0.03$)
Log(Qual)	$.939 \times 10^{-1}$ ($p < 0.28$)	$.459 \times 10^{-1}$ ($p < 0.66$)
Qual^2	$-.138 \times 10^{-4}$ ($p < 0.39$)	$-.1 \times 10^{-6}$ ($p < 0.53$)
Task ease	$.34 \times 10^{-5}$ ($p < 0.53$)	$.14 \times 10^{-4}$ ($p < 0.01$)
R^2	.20	.52
F(5,139)	4.63 ($p < 0.001$)	
F(5,43)		11.68 ($p < 0.001$)

NOTE: *p*-values in parentheses.

Staelin's study.[2] The smallest level was designed to reflect subjects whose weight measures were largely constant over time and, hence, reflected a level of noise undoubtedly smaller than in their experiment. Conversely, the covariation pattern was designed to be more extreme and characterized a case with almost no error in assessing unimportant attributes but a high error (50 percent) in judging the most important, with a modest positive correlation. In each choice problem, the true weight estimates were equal across attributes, formally being given by the designated sum of the weights in that condition divided by the given number of attributes. Attribute values for each option were randomly drawn from independent uniform distributions defined over a [1–100] range.

Our central interest was the extent to which effects of quantity and quality would be sustained in each covariation condition after controlling for task ease, as per Keller and Staelin. Mirroring Keller and Staelin, task ease was defined as the absolute difference in overall predicted utility values within each paired comparison, using the true weight sets for each condition (i.e., the weights before random noise was applied). We then regressed the proportion of correct predictions observed for each quantity, quality, and noise combination against the variables considered in Keller and Staelin's regression analysis: the log of information quantity (Log(N)), the log of total quality (Log(Qual)), total quality squared (TQ^2), and task ease (here the average ease of the tasks within conditions).[3] The results of this analysis are summarized in the Table.

The first central finding is a significant negative Log(N) effect even when task ease is controlled for and errors positively covary with attribute weight—a finding directly mirroring Keller and Staelin's results under their assumptions. Although introducing a positive correlation between error and weight did somewhat reduce the effect of information quantity, as predicted, it did not fully eliminate it. In addition, although we were not able to support a significant quality effect, in both error conditions the direction follows that reported by Keller and Staelin: Log(Qual) was positively associated with accuracy and Qual^2 negatively,[4] implying diminishing positive returns to increases in information quality, as in Keller and Staelin.

Why did the correlation manipulation fail to mitigate the quantity and quality effects? A positive correlation between the size of the random error and attribute weight would fully eliminate a negative quantity effect only under the rather extreme condition where error variance was a constant positive multiple of the weight value (i.e., $\mathrm{VAR}(\epsilon_w) = \alpha w_k$, $\alpha > 0$). We imposed a positive correlation between weight value and error in the simulation, but the relationship was nonlinear. In similar cases where constant proportionality does not hold, the extent to which a positive correlation will negate quantity effects is ultimately an empirical issue. The result shows that for the one covariation pattern modeled in the experiment, this weakening effect was negligible. Hence, even if there was a positive correlation between weight and weight error in Keller and Staelin's task, if it was less extreme than that which we modeled here, it would likely be insufficient to negate an artifactual negative effect of information quantity.

Taken together, these simulation results show how information quantity and quality independently affect decision accuracy and bear a striking resemblance to results reported by Keller and Staelin. The central difference from Keller and Staelin's analysis, however, is that these results were generated without assuming any decrement in the information processing ability of consumers due to information overload.

DISCUSSION

There is little doubt that increases in information do not have uniformly positive effects on consumer decisions. For example, Payne (1976) and others have noted that increases in information often induce the use of heuristic evaluation policies that do not

[2]Keller and Staelin did not measure the error rates that existed for attribute weight and score judgments. They did, however, report that there was rather extensive between-session variation in weight judgments, with an average between-session rank correlation of 0.57.

[3]We also conducted this regression on the limited fraction of the design that Keller and Staelin considered in their study (1987, p. 206, Table 2). The results of these analyses approximated those for the full factorial that we report.

[4]A possible reason for our failure to establish significant effects for both information quality and task ease in the independent error condition was the extremely high (−0.87) correlation that existed between these two constructs. Keller and Staelin indicated in a personal communication that collinearity was a less serious problem in their analyses because they hand-screened choices for clearly dominated options and conducted their analyses at a disaggregate level, where task ease would have greater variance.

consider simultaneously the value of all options on all attributes. Similarly, Scammon (1977) has noted that increases in information decreases individuals' abilities to recall product attributes. What is less clear, however, is that increases in information can actually be harmful in the sense that they cause consumers to make different choices than would have been made with broader information processing powers. This turns out to be a difficult proposition to support simply because there will always be measurement error when using models to define the optimal decision. If the optimal decision is computed from a set of independently derived attribute weight measures, the measurement error will often increase naturally as a function of information quantity or the number of attributes under study.

In our view, Keller and Staelin's conclusions about how information quantity affects decision efficiency must be tempered in light of this potential confound. There is no question that decision accuracy decreased as information quantity increased, but less certain is the interpretation of this. That error and task structure would produce such a decrease suggests that choice errors in the experiment were those of the predictors, not the subjects.

Error in measurement represents a formidable, albeit not the worst, obstacle to using preferences inferred from models to study decision quality. Studies of information overload traditionally have employed one method of measurement to calibrate a normative choice model (e.g., Keller and Staelin used subjects' judgments of the independent importance of attributes) and another for judging errors (using choices from sets). Unfortunately, there is strong evidence that measures of preference often are inseparable from the method used to elicit them. Starting in the late 1960s, Lichtenstein and Slovic (1971) produced a number of demonstrations that preferences among simple gambles implied by judgments of monetary value often differ from the choices actually made. Likewise, Tversky, Sattath, and Slovic (1988) and Fischer and Hawkins (1987) have offered demonstrations that similar reversals as a function of response mode can be obtained even in riskless choice.

The implication of this for measuring decision overload using inferred choice errors seems clear. Because quite different cognitive processes may be used to generate different types of preference judgments, it will never be obvious which set of measures reflects most accurately the subjects' true normative preferences. Identifying overload requires a relatively precise and stable representation of underlying preferences that is difficult to obtain in practice.

Although we have focused our discussion on the problems of inferred choice in the study of information overload, we should stress that the same concerns apply more broadly to all research in which the quality of decision making is examined. Thus, work that examines the effectiveness of consumer decision aids, or the effect of expertise or familiarity on decision performance, faces difficulty in defining a good decision. It is our view that the only means by which decision inefficiencies can be studied without ambiguity is through demonstrations of inconsistencies that are relatively insensitive to measurement error, such as subjects' picking dominated options as defined by an invariant measurement procedure. Unfortunately, this is a standard of rigor applied too infrequently in studies of decision quality.

By offering an alternative view of Keller and Staelin's data, our intent was not to downplay their contribution but to clarify it by stressing the inherent difficulties in testing for the effects of information overload. This difficulty is shared by all researchers in the area. The study of information overload, perhaps more than any other in the consumer research literature, is one that historically has progressed with the aid of academic exchanges such as this one. Our hope is that this communication offers a similar contribution.

[*Received November 1987. Revised June 1988.*]

REFERENCES

Fisher, Gregory and Scott Hawkins (1987), "Riskless Preference Reversals and Information Processing Strategies," working paper, Department of Social and Decision Sciences, Carnegie-Mellon University, Pittsburgh, PA 15213.

Keller, Kevin L. and Richard Staelin (1987), "Effects of Quality and Quantity on Decision Effectiveness," *Journal of Consumer Research,* 14 (September), 200–213.

Lichtenstein, Sarah and Paul Slovic (1971), "Reversal of Preferences Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology,* 89 (July), 46–55.

Payne, John W. (1976), "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance,* 16 (August), 366–387.

Scammon, Debra L. (1977), "Information Load and Consumers," *Journal of Consumer Research* 4 (December), 148–155.

Tversky, Amos, Shmuel Sattath, and Paul Slovic (1988), "Contingent Weighting in Judgment and Choice," *Psychological Review,* 95 (July), 371–384.