COMMENTS ON "HOW MANY OPTIONS? BEHAVIORAL RESPONSES TO TWO VERSUS FIVE ALTERNATIVES PER CHOICE"

BY MARTIN MEISSNER, HARMEN OPPEWAL, AND JOEL HUBER

JOHN R. HAUSER

MIT SLOAN SCHOOL OF MANAGEMENT
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AN INSIGHTFUL SCIENTIFIC PAPER

I would like to congratulate the authors on a very interesting study and presentation. Intuitively, we all want to design a conjoint study with tasks that best approximate the marketplace. But which tasks and which formats best approximate the marketplace? I am glad to see that many papers in this conference addressed this question. There are papers on number of alternatives, whether small screens compromise choice tasks, and whether dual-response formats are important. Other papers address whether or not we compromise the integrity of our analyses if, rather than using profession panels, samples are based on Amazon's Mechanical Turk Panel. Earlier in this session, my colleague, Felix Eggers, addressed the realism of the stimuli, incentive alignment, training videos, and ceteris paribus instructions.

These issues of "craft" are extremely important. By definition, better craft leads to more-accurate relative partworths and a greater scale factor (precision). Craft matters because accuracy and precision matter. Managerial recommendations about which attribute levels to include in new products depend on accuracy and precision, as do simulations of marketplace response (such as predicted equilibrium prices and profits).

We should not neglect the development and evaluation of experimental and statistical methods, because these methods also impact accuracy and precision. Developments such as ACBC, non-compensatory estimates, empirical Bayes, and machine-learning methods all have the potential to improve managerial recommendations. These developments are all extremely important, but we cannot neglect craft. Craft matters!

HOW MANY OPTIONS?

The authors ask: "How many options are best for choice sets in a choice-based conjoint analysis (CBC) design?" They present evidence that practical applications of CBC vary from two options to five options. Over the past four years, two options have become rare with three options emerging as the dominant design. Five options were once common, but by 2015 have become rare.

The issue is clear, but the answer is not. More options provide more data per question in the form of "constraints." For two options we know that one profile is preferred to the other. For five

options we know that one profile is preferred to four other profiles. On the other hand, as the number of profiles per choice set increases, the task becomes more demanding. For a given amount of respondent effort, respondents might be able to make more two-option choices than five-option choices. Not only might the difficulty of the task for five options lead to respondent wear-out (for a fixed number of choices), but the behavioral literature suggests that respondents might process five options differently than two options (e.g., Payne, Bettman, and Johnson 1988).

The authors design is inspired. By comparing the two extremes found in practice, they are able to analyze the differences in the way consumers process information. Practically, the authors identify that two options are better than five options. This encourages researchers to search for the optimal—is it two, three, or four options—and to determine if the optimal number depends on context. The authors suggest a context, with which I agree. If consumers use a consider-then-choose heuristic, and there is strong evidence that consumers do so, then five options might be better than two for modeling the consideration decision while two options might be better than five for modeling the choice-given-consideration decision.

The authors were kind enough to provide me with data to determine whether two options or five options leads to more-precise estimates ($\gamma^{estimation}$) as defined by Eggers, Hauser, and Selove (2016) in this volume. For beach-hotel-room choice, two options (pairs) do better than five options (quints).

- Precision is 64% larger.
    - The sum of importances is 36.9 for pairs.
    - The sum of importances is 22.5 for quints.
- Percent of uncertainty explained (U^2) is 31% larger.
    - U^2 = 0.90 for pairs.
    - U^2 = 0.69 for quints.
- Relative hit rates are 27% larger.
    - The relative hit rate is 56% for pairs.
    - The relative hit rate is 44% for quints.

These numbers convince me that, for product categories similar to the authors' (beach hotel rooms), I would prefer to use pairs to model choices (given consideration) rather than quints. In this case, the better craft is two options. (I await with excitement, comparisons to triples and quads.)

EVIDENCE OF NON-COMPENSATORY PROCESSING

Having determined the best craft for their product category, the authors addressed the scientific question of "why?". In a study of high scientific craft, the authors use eye-tracking to compare how respondents evaluate pairs and quints.

The results are fascinating. For pairs, the respondents tended to compare attribute levels between products. For quints, respondents began with a comparison of attribute levels among

products, but respondents quickly zeroed in on one or two "finalists." After zeroing in on finalists, the respondents did more vertical, within product, processing.

There is evidence in the data that quints push respondents toward non-compensatory heuristic processing. In particular, we see evidence consistent with non-compensatory decision rules. In particular, if the respondent uses a non-compensatory decision rule for quints:[1]

- We expect that the variance of the estimated importances increases with quints. It does.
- We expect to see a drop in perceived processing effort with quints, because respondents are using heuristic shortcuts. Perceived difficulty drops.
- We expect to see a steeper decrease in effort (eye fixations) over tasks as respondents develop fast heuristics. There is a steeper decline.

While this evidence does not prove that consumers are using non-compensatory processes for quints, it is consistent with the hypothesis. If we combine these hints with the authors' more-detailed eye-tracking analysis, there is strong evidence that consumers are using non-compensatory decision processes for quints. This makes sense. The literature suggests that consumers use heuristic rules for consideration and compensatory rules for choice within the consideration set. It is reasonable to hypothesize that the choice of format depends upon the decision being modeled, something that can be assessed further with measures that approximate the external validity of the consumers' decisions.

The authors have succeeded on both a practical and a theoretical level. They deserve the praise they have received.

REFERENCES

Eggers, Felix, John R. Hauser, Matthew Selove (2016), "The Effects of Incentive Alignment, Realistic Images, Video Instructions, And Ceteris Paribus Instructions on Willingness to Pay and Price Equilibria," *Proceedings of the 19th Sawtooth Software Conference,* Park City, UT, September 26-20.

Meissner, Martin, Harmen Oppewal, and Joel Huber (2016), "How Many Options? Behavioral Responses to Two versus Five Alternatives per Choice," *Proceedings of the 19th Sawtooth Software Conference,* Park City, UT, September 26-20.

Payne, John W., James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision Maker*, (Cambridge UK: Cambridge University Press).

[1] A decision rule is compensatory if high levels on one or more attributes can compensate for a low level of one attribute. A non-compensatory decision rule is a rule that is not compensatory. Typical non-compensatory rules are conjunctive (consider a product if it has certain must-have attribute levels) or lexicographic (rank on a single attribute, continuing to the next only if the products are tied on the most-important attribute—as a dictionary ranks words in alphabetical [lexicographic] order). Additive representations, as in conjoint analysis, can be compensatory or non-compensatory. For example, if the partworths of four dummy-coded attributes are 8, 4, 2, and 1, then the decision rule is lexicographic.