EDITORIAL

Comment on: Taking the Shortcut: Simplifying Heuristics in Discrete Choice Experiments

F. Reed Johnson[1]

I recently purchased a new, all-electric, red, crossover SUV. I was quite aware that I was using several simplifying heuristics in making a consequential purchase. Moreover, my wife observed that my decision clearly was not rational. My decision appeared irrational to her because we disagreed on both the utility of various automobile characteristics and the relevant budget constraint. In addition, I had clearly defined trade-off preferences that were non-compensatory in several dimensions and involved specific restrictions on the consideration set. Nevertheless, even my wife would not suggest that my using simplifying heuristics made my purchase ineligible for inclusion in the US gross domestic product.

In 1977, George Stigler and Gary Becker published a classic article in the *American Economic Review* titled "De Gustibus Non Est Disputandum" (in matters of taste, there can be no dispute) [1]. Stigler and Becker contrasted two approaches to thinking about tastes. One view is that there is no role for economic analysis when variations in economic decisions are judged to be a result of inscrutable, often capricious, variations in tastes. The problem then must be turned over to whomever studies taste formation, presumably psychologists. Stigler and Becker argued that a far more analytically productive approach is to assume that tastes generally are stable among people and over time, and economists should instead focus on a generalized calculus of utility maximization constrained by variable resource availability and prices. Becker demonstrated the power of this approach by opening new fields of economic research in such areas as marriage and reproductive decisions, household decision making, addiction, crime, and political markets.

The power of Stigler and Becker's theory of consumer choice derived from the simple assumption that people act as if they were maximizing a defined objective or utility function subject to money, time, and information constraints. It did not require examining actual thought processes or emotional states of decision makers, or asking them to explain how they reached a particular decision. All that mattered was that they acted *as if* they were rational in this specific purposeful sense. Unfortunately, behavioral economists, who often are not formally trained in economics, have exploited the marketability of the idea that economics is bad psychology. In fact, conventional economists are entirely agnostic about what arguments should be included in utility functions. Objective functions could include concerns for the utility of other people (benevolent or malevolent), value of uniformity or variety, religious belief or unbelief, or any other factors without judging the appropriateness, social acceptability, or even reasonableness of what people care about.

The article authored by Veldwijk, Swait, and deBekker-Grob was an invited submission to *The Patient* offering analytical guidance on identifying and handling respondents' use of simplifying heuristics to answer discrete-choice experiment (DCE) questions. The authors are leading stated-preference researchers, which is reflected in their comprehensive catalog of heuristic types and excellent advice on analysis. It naturally is frustrating when survey participants do not give us data amenable to a straightforward analysis. However, I have concerns about the authors' assertion that heuristics violate one or more assumptions of McFadden's random-utility theory (RUT) [2]. When McFadden appended a random component to ordinary utility functions to account for measurement errors, he constructed a framework that facilitated the statistical analysis of choice data but did not alter any conventions about the properties of utility functions. Thus, everything that Stigler and Becker thought about utility was still part of that framework. One key aspect of RUT, however, is that the random component must truly be

✉ F. Reed Johnson
  reed.johnson@duke.edu

1 Department of Population Health Sciences, Duke Clinical Research Institute, Duke University, 300 West Morgan Street, NC 27701, NC, Durham, USA

random and uncorrelated with variables in the deterministic term of the utility function.

Before concluding that decision heuristics in choice data compromise the assumption of rational behavior and turning to psychologists for help, Stigler and Becker would ask: are you sure you included all the attributes and attribute ranges relevant to participants' utility in the study design and have you accounted for perceived time, effort, and information constraints? Given that the RUT error structure accounts only for measurement error, it is fair ask to what extent estimation problems stem from systematic variation in the error term not accounted for in the deterministic term in the specified DCE utility function.

As early as 1947, Herbert Simon described satisficing as utility maximization that appears inefficient in strictly monetary terms, but not when accounting for nonmonetary costs of information acquisition, decision effort, and institutional constraints [3]. Subsequently, economists accommodated a wide range of potential altruistic, cultural, and normative arguments in utility functions, as well as divergence from conventional expected-utility assumptions using prospect-theoretic probability weighting and rank-dependent utility within a generalized RUT model. Survey researchers also have evaluated the effect of satisficing behavior in survey responses [4].

We all frequently make decisions that violate a narrow definition of RUT because they involve rational satisficing. In many cases, the incremental time and effort required to reduce the likelihood of making relatively inconsequential, and usually reversible, mistakes just are not justified. Hence, for many people taking a DCE survey, it is quite rational not to devote much time and effort to process the information required and to explore the relative importance of multiple attributes, each with multiple levels. For them, the perceived benefit of carefully answering such survey questions does not justify the associated effort. The bad-data problem in DCE studies thus in large part could be a result of rational behavior that leads to optimally inattentive responses [5].

The authors acknowledge that "It has been shown that even when ignoring information, certain heuristics do allow respondents to arrive at an optimal decision without allocating full effort." Nevertheless, "although heuristics help individuals solve complex decision-making in real life, their use may result in choices made in partial or total conflict with [strict] RUT assumptions." Random-utility theory decision makers are assumed to choose utility-maximizing alternatives from a fixed choice set, where utility maximization is consistent with various logic and consistency conditions. However, it is hard to see why DCE participants have to care about each of a small subset of factors varying over particular ranges. These attribute ranges are used to construct artificially constrained choice sets with known statistical properties, but having no particular relationship to actual or even realistic choice sets. Hence, violations of strict RUT assumptions could well be attributable to study designs that fail to incorporate preference-relevant trade-off options for many participants.

The authors identify seven DCE-specific heuristics (attribute and level non-attendance, choice-set screening, elimination by aspect, lexicographic preferences, attribute dominance, and satisficing). Much of the behavior described is consistent with a mismatch between the study design and attributes and attribute levels that actually are utility relevant to respondents. The authors suggest various approaches to identifying and controlling for alternative data-generation processes, but these data-generation processes could signal problems with study designs and could be informative about respondents' actual preferences.

The challenge of analyzing choice data is that we do not have good methods for discriminating between heuristics resulting from inattention and heuristics that are consistent with well-defined preferences. The best example is attribute dominance. An attribute could well be strongly preferred but the study design never shows levels of less-preferred attributes large enough to induce trading away from the better level of the dominant attribute. While some researchers advocate using overlapping levels for potentially dominant attributes to force participants to consider less-important attributes, this approach does not necessarily ensure valid trade-off data. Participants in a vaccine study who chose an opt-out alternative then were asked a forced-choice evaluation of the two opt-in alternatives [6]. All participants chose the vaccine with the lower cost. When we tried to make them buy something they did not want, they logically minimized the cost of the forced choice rather than providing information about their trade-off preferences among the attributes of interest.

The authors advocate a good survey-instrument design to effectively convey relevant information and to minimize cognitive burden. This of course is a valid and important recommendation. However, strategies to influence participants' perception of the net benefits of allocating effort to the choice tasks could be even more important. Stated-preference research in environmental economics has documented the importance of "consequentiality" in obtaining valid choice data [7, 8]. "Cheap talk" sometimes has been shown to improve data quality by encouraging participants to think of their role as being part of a research team to help researchers, physicians, and government officials expand treatment options for their condition [9]. In fact, compared to many environmental applications, DCE applications in health generally produce quite good data. Patients often value researchers' interest in their disease experience and in participating in studies aimed at improving available treatments. When that interest more than compensates for the time and effort required to take

a survey, we can get data that reflect good attention to the information provided and careful assessment of the trade-offs in the study design.

If heuristics are strategies participants use to simplify choice tasks to make them easier to answer, participants also could employ strategies to compensate for missing information or to resolve inherent contradictions, resulting in choice questions that make no sense. For example, in a study of epilepsy treatments, participants were asked to evaluate seizure reductions of 25%, 50%, and 75% relative to the number of seizures they experienced in the previous 3 months despite the fact that half of the participants had not experienced any seizures during that period [10]. Nevertheless, the categorical efficacy parameters were logically ordered and well spaced, with tight confidence intervals. Faced with an illogical question, participants evidently decided the researchers must have intended some 3-month period in which they had enough seizures to make a 75% reduction meaningful. Unfortunately, interpreting the results was difficult without information about the assumed baseline levels.

Researchers often resist pivot designs such as that used in the epilepsy study that allow the reference condition to be endogenous to each respondent. The gain in personal relevance and engagement is offset by losing some control over the experiment and the possibility that responses will generate insufficient information about trade-offs in the policy-relevant space. Similarly, attribute ranges could be endogenous to avoid showing strongly risk-averse respondents large side-effect probabilities and strongly risk-tolerant respondents' small side-effect probabilities they will ignore. Market researchers often use self-explicated questions at the beginning of a survey to tailor consideration sets and attribute ranges to help define participant-relevant trade-offs. Showing different respondents different attributes and attribute ranges reduces the statistical power of the experimental design, but could reduce measurement error induced by experimental designs that encourage non-attendance and screening heuristics.

While preference researchers tend to see opportunities for hitting almost every medical decision-making nail with a DCE hammer, the authors' survey of challenges of eliciting and analyzing choice data suggests there is room for some humility about the limitations of these methods. There are questions that could be too complicated for distilling into a tractable DCE instrument. An unusual lessons-learned assessment of a study of whole-genome sequencing preferences concluded that the inherent dimensionality of the problem defied efforts to apply the best available DCE methods [11]. Nevertheless, people rationally using decision heuristics commonly manage to make whole-genome sequencing choices that resolve trade-offs between the

likely value of genetic information and monetary and non-monetary costs.

Veldwijk, Swait, and deBekker-Grob have given us excellent guidance on the challenges of identifying response patterns that complicate the analysis of choice data. Having identified such patterns, there are opportunities for additional research to devise patient-centric DCE study designs that relieve methodological restrictions that contribute to survey participants employing heuristics that mask their actual trade-off preferences. An alternative to searching for respondents who comply with strict RUT assumptions could be searching for DCE strategies that comply with people maximizing well-defined utility functions subject to money, time, and information constraints.

Declarations

Conflicts of Interest Not applicable.

Ethics Approval Not applicable.

Consent to Participate and Consent for Publication Not applicable.

Availability of Data and Material Not applicable.

Code Availability Not applicable.

References

1. Stigler GJ, Becker GS. De gustibus non est disputandum. Am Econ Rev. 1977;67(2):76–90.
2. Veldwijk L, et al. Taking the shortcut: simplifying heuristics in discrete choice experiment. Patient. 2023. https://doi.org/10.1007/s40271-023-00625-y
3. Simon HA. Administrative behavior: a study of decision-making processes in administrative organization. 1st ed. New York: Macmillan; 1947.
4. Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. Med Decis Making. 2011;31(3):469–80.
5. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cogn Psychol. 1991;5(3):213–36. https://doi.org/10.1002/acp.2350050305.
6. Brown DS, Johnson FR, Poulos C, Gonzalez JM, Messonnier MM. Mother-daughter preference conflicts and willingness-to-pay for HPV vaccines. In: Presented at the international health economics association conference. Beijing, China; 2009.
7. Welling M, Zawojska E, Sagebiel J. Information, consequentiality and credibility in stated preference surveys: a choice experiment on climate adaptation. Environ Res Econ. 2022;82(1):257–63. https://doi.org/10.1007/s10640-022-00675-0.

8. Carson RT, Groves T, List JA. Consequentiality: a theoretical and experimental exploration of a single binary choice. J Assoc Environ Res Econ. 2014;1:171–207.

9. Ozdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J Health Econ. 2009;28(4):894–901. https://doi.org/10.1016/j.jhealeco.2009.04.004.

10. Manjunath R, Yang JC, Ettinger AB. Patients' preferences for treatment outcomes of add-on antiepileptic drugs: a conjoint analysis. Epilepsy Behav. 2012;24(4):474–9. https://doi.org/10.1016/j.yebeh.2012.05.020S1525-5050(12)00404-0.

11. Marshall DA, Gonzalez JM, MacDonald KV, Johnson FR. Estimating preferences for complex health technologies: lessons learned and implications for personalized medicine. Value Health. 2017;20(1):32–9. https://doi.org/10.1016/j.jval.2016.08.737.