Annual Review of Resource Economics

Methodological Advances in Food Choice Experiments and Modeling: Current Practices, Challenges, and Future Research Directions

Vincenzina Caputo[1] and Riccardo Scarpa[2]

[1]Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, Michigan, USA; email: vcaputo@msu.edu

[2]Durham University Business School, Durham University, Durham, United Kingdom; email: riccardo.scarpa@durham.ac.uk

Annu. Rev. Resour. Econ. 2022. 14:63–90

The *Annual Review of Resource Economics* is online at resource.annualreviews.org

https://doi.org/10.1146/annurev-resource-111820-023242

JEL codes: D01, D12, C9, C25, C49, C83

Keywords

discrete choice experiment, food choice, methodological advances, choice experiment designs, discrete choice models

Abstract

In recent decades, discrete choice experiment research applied to food choices has grown rapidly. Empirical applications include investigations of consumer preferences and demand for various food attributes, labeling programs, novel products and applications, and new food technologies. Methodological contributions include advances in the form of new theories, elicitation methods, and modeling. This study focuses on the latter and (*a*) reviews recent methodological contributions in the food choice experiment literature, (*b*) examines existing knowledge gaps, and (*c*) discusses possible future research directions.

1. INTRODUCTION

In recent decades, discrete choice experiment (DCE) research has grown rapidly in applied economics, particularly so in the subfields of environmental, agricultural, health, and transportation research. According to the Web of Science, in 2020 alone, 727 DCE articles were published, implying a more than 2,800% increase since 2000 (**Figure 1**).

The popularity of DCEs has also grown in food preference studies owing to the ability to produce easy-to-interpret results in terms of probabilistic models of inferential choice. This stems from the core advantage of DCEs over other nonmarket valuation methods (Lusk & Schroeder 2004) such as contingent valuation (Hanley et al. 2001) and experimental auctions (Lusk & Shogren 2007): DCEs are consistent with both random utility theory and Lancaster's theory of value decomposition in consumer demand, allowing researchers to simultaneously evaluate multiple food attributes and products. In addition, compared to revealed preference data, DCEs allow researchers to control and manipulate information provided to participants, to include, for example, some features that may not yet be available in markets. Prices are also determined exogenously by researchers, enabling them to avoid endogeneity issues. Further, their high degree of external validity and predictive power make food DCEs a suitable substitute for and complement to revealed preference data, such as scanner data (Chang et al. 2009, Brooks & Lusk 2010).

Empirical applications of food DCEs include investigations of consumer preferences and potential demand for various food quality attributes (Darby et al. 2008, Scarpa et al. 2008), labeling programs (Loureiro & Lotade 2005; Onozaka & McFadden 2011; Aprile et al. 2012; Caputo et al. 2013a,b), novel products (Van Loo et al. 2020), and new food technologies (Caputo 2020, Kilders & Caputo 2021), as well as ex ante food policy evaluations (Papoutsi et al. 2015, Ahn & Lusk 2021). Yet, the external validity and reliability of findings from food DCEs depend on whether experiments can accurately predict actual choice behavior in the market from hypothetical choices obtained in experimental conditions. As such, the food DCE literature has also investigated (*a*) which experimental designs can best capture actual food purchasing situations while remaining practical and affordable and (*b*) what behavioral assumptions, statistical methods, and hypothetical bias correction methods can improve the predictive accuracy of choice models in food demand analysis.

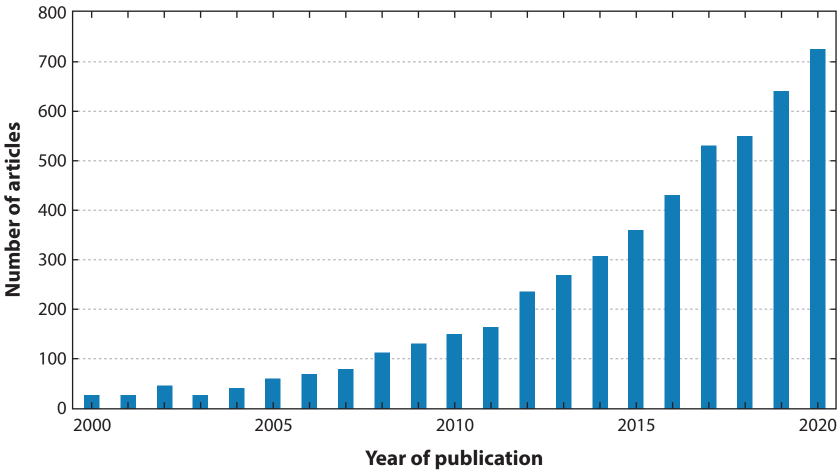

Figure 1

Number of publications and their growth rate on the topic of discrete choice experiments in applied economics from 2000 to 2020. Based on data from the Web of Science (query executed in September 2021).

Collectively, this methodological work has led to conceptually novel choice models (Caputo et al. 2018a, 2020), new elicitation methods (Scarpa et al. 2013, 2021a; Malone & Lusk 2018b; Caputo & Lusk 2020; Dennis et al. 2021), and practical recommendations (Balcombe et al. 2009, Caputo et al. 2018b) that have improved food DCEs by showing choice determinants including experimental designs, behavioral assumptions, and modeling, among others. Ignoring these factors can lead to biased estimates of policy-relevant quantities and possibly invalid research findings.

In this article we (*a*) review recent methodological contributions in the food DCE literature, (*b*) attempt to identify existing knowledge gaps, and (*c*) propose and discuss possible future research directions. We hope to provide both a comprehensive guide to researchers interested in using the methodology and information for interested stakeholders on how to evaluate existing and future results in publications in this area of food research.

The review is organized as follows. In Section 2 we discuss external validity in food DCEs, which is followed by an evaluation and review of current best practices for the design of DCEs in Section 3. Section 4 covers the actual modeling of DCE data. Section 5 discusses the behavioral factors needed to be considered in DCEs, while Section 6 focuses on the relevance of DCEs for food policy evaluations. Section 7 concludes.

2. EXTERNAL VALIDITY OF FOOD DISCRETE CHOICE EXPERIMENTS

The validity and reliability of findings from food DCEs depend on whether experiments can reproduce the essential traits of real-world situations and thereby capture real-world behavior. Unlike other nonmarket valuation methods such as the contingent valuation of public goods (Vossler et al. 2003) and experimental auctions (Lusk & Shogren 2007),[1] DCEs present subjects with more familiar choice contexts that mirror more closely what people experience in real-world food shopping situations. During the experiment, subjects are presented with a hypothetical market mechanism that mirrors what they would experience in a real food choice: The seller offers various products at different posted prices and consumers choose which products to buy, if any. This choice is akin to when consumers select their favorite product alternative from a supermarket shelf or a market stall displaying various alternatives. Nevertheless, questions regarding the external validity of DCEs remain. Assessing whether inferences obtainable from food DCEs are externally valid remains a crucial issue when proposing the use of this method to inform agribusiness decisions and policy design/implementation. Field studies, like those carried out for the validation of contingent valuation in the early 2000s, for example, by Vossler et al. (2003), are still too few and often inconclusive.

External validity can be defined "as the ability to generalize the relationships found in a study to other persons, times, and settings" (Roe & Just 2009, p. 1267). In the DCE literature, researchers have used two approaches to study the extent to which stated food choices from DCEs can be generalized to wider populations or to other choice contexts. The first approach combines stated food choice data and revealed preference data to determine (*a*) whether survey-based hypothetical choices are consistent with people's revealed preferences, and (*b*) whether the accuracy of welfare estimates, and their preference structure, can be improved by doing so. The seminal study by Brooks & Lusk (2010) combined DCE data with scanner data to determine consumer demand for organic and cloned milk. Results suggest that DCE data once merged with revealed preference

[1]Contingent valuations and experimental auctions exemplify experimental settings that consumers are unlikely to experience in actual shopping situations [e.g., bidding for a good or directly indicating the willingness to pay (WTP) for it].

data can substantially improve the out-of-sample prediction performance of the model compared to either DCE or revealed preference data used in isolation. Most notably, their results show that models estimated on revealed preference data yield incorrect demand forecasts when used to predict demand for products with novel features. The second approach compares data from hypothetical and real choice experiments to assess which type better mirrors actual retail shopping behavior. For example, to predict actual retail shopping behavior in three different product categories (ground beef, wheat flour, and dishwashing liquid), Chang et al. (2009) compared three elicitation methods: hypothetical choices, non-hypothetical choices, and non-hypothetical rankings. They found that the predicted market shares correspond well with the actual market shares across the three elicitation formats, although the non-hypothetical choices were found to be a better approximation of true preferences than hypothetical choices, likely due to hypothetical bias.

Taken together, these initial results suggested that model estimates from stated food choices may hold substantial validity and could be used to predict food choices in actual food markets and to complement and improve the predictive power of estimates from revealed preference data. Yet, the external validity of DCEs remains relatively unexplored within the food choice literature and beyond, especially over time and over places. A key issue when combining hypothetical choices from DCEs and real food choices from revealed preference data is to consider the nature and composition of the two data sets (see Hensher et al. 1998 for a detailed discussion). For example, factors such as different temporal horizons, data structure, choice set compositions, and biases from subjective recalls are of concern. Likewise, when comparing DCEs and scanner data, researchers need to make several assumptions regarding the (un)availability of product alternatives, observed price distributions, and nonpurchases. Further concerns related to scanner data that researchers need to account for include price and expenditure endogeneity (for a review, see Caputo & Just 2022), and this even before combining or comparing such data with stated choices.

In the light of this, recent research efforts have focused on reevaluating the approaches researchers should follow to best assess the external validity of DCEs. Those used so far have been inherently backward-looking; they assess whether consumers display the same behavior in real and hypothetical markets. Lancsar & Swait (2014) propose a more holistic forward-looking approach, whereby external validity is closely linked to process validity, rather than only being a comparison of final choice outcomes between stated choices and revealed preference data. The authors suggest the use of qualitative research to inform DCE design and implementation. They also outline three key considerations as to why models estimated from DCEs may fail external validity tests. These include (*a*) the context in which the study is undertaken, (*b*) the realism of the assumptions underpinning the model and validity tests, and (*c*) the internal validity of the DCE. In this regard, induced value experiments are emerging as a valid method to assess the validity of homegrown[2] DCEs for both public (Collins & Vossler 2009) and private (Luchini & Watson 2014) goods. As argued by Collins & Vossler (2009), the use of induced value DCEs allows researchers to test whether observed choices align with induced preferences and whether estimated welfare measures such as WTP are equal to induced values. Indeed, recent studies have used induced value food DCEs to test the assumptions underlying different empirical distributions of discrete choice models (Bazzani et al. 2018, Carson et al. 2020) and the incentive compatibility[3] of DCEs (Cerroni et al. 2019). Further studies using this validation process are recommended.

---

[2]Homegrown DCEs focus on private goods, commodities, or policy.
[3]An elicitation method is considered incentive compatible if it provides incentives for individuals to truthfully reveal their true preferences/values and imposes a cost for inaccurate value revelation.

2.1. Hypothetical Bias in Food Choice Experiments

Despite past studies having shown that DCEs display substantial external validity, and new evidence in support of stated preference data [for the specific literature in environmental good valuation, see the review by Kling et al. (2012) and the proposed calibration of stated/revealed preference values by Murphy et al. (2005); for general guidance, refer to Johnston et al. (2017) and Mariel et al. (2021)], the presence of potential hypothetical bias,[4] although thought to be stronger in the valuation of environmental goods than in private goods because of the higher degree of familiarity in the latter, remains a major concern across all stated choice data studies. Yet, conclusions drawn from the DCE literature in food preference analysis are less supportive of hypothetical bias and often contradictory. Lusk & Schroeder (2004) compared marginal and total WTP estimates from hypothetical and non-hypothetical responses to DCE questions on beef steak selection. The authors found that the estimated average total WTP for steaks in the hypothetical setting was approximately 1.2 times that in the non-hypothetical setting. However, no statistical differences were found between estimates of averages of marginal WTP across hypothetical and non-hypothetical experiments. Carlsson & Martinsson (2001) and Cameron et al. (2002) also found no statistical differences in marginal WTP estimates between real and hypothetical DCE settings in environmental economics.

Nevertheless, concerns about hypothetical bias have led several authors to introduce various ex post and ex ante bias-reducing methods. Ex post methods include calibration techniques using data obtained from hypothetical and non-hypothetical experiments (see, e.g., Fox et al. 1998, List & Gallet 2001) and the use of certainty follow-up questions (see, e.g., Ready et al. 2010). Ex ante methods include scripts (or tasks) that survey respondents are asked to read (or complete) prior to the choice exercise, which act as psychological nudges to avoid hypothetical bias. These include cheap talk scripts, solemn oaths, honesty priming, consequentiality, opt-out reminders, and inferred methods. Most of these methods were initially developed for contingent valuation studies of public goods and subsequentially applied to DCEs. **Table 1** outlines seminal studies and the first fields in which these correction methods were applied.

Since the introduction of the first ex ante hypothetical bias mitigation methods, a large and growing body of food choice literature has investigated their efficacy and performance. One strand of the literature compares welfare estimates from non-hypothetical and hypothetical stated food choices with and without the use of a single ex ante hypothetical bias correction method. This literature has drawn contradictory findings across correction methods. For example, Moser et al. (2014) studied apple choices by comparing hypothetical choices with and without ex ante bias-reduction techniques (cheap talk scripts) with real choices. They found significantly higher WTP estimates for hypothetical choices and a limited performance of cheap talk in attenuating this bias. Alemu & Olsen (2018) conducted hypothetical and non-hypothetical field DCEs in Kenya on choices between attribute quality–differentiated buns. They found a severe hypothetical bias and that the opt-out reminder effectively reduces or eliminates hypothetical bias, while also attenuating or removing negative opt-out effects. Other studies used purely hypothetical experiments with one of the correction methods.[5] Inconclusive findings are often reported in these studies. For example, Lewis et al. (2016) found that the use of a consequentiality script did not affect welfare estimates, while Carlsson et al. (2005) found that cheap talk significantly reduces welfare estimates.

[4]Hypothetical bias is defined as the difference between the amount of money people state they are willing to pay in hypothetical experiments and what they would actually pay in real purchase situations.

[5]The studies comparing ex ante correction bias techniques using hypothetical settings assume that the lower WTP estimates are the ones less affected by hypothetical bias.

Table 1 The seminal publications introducing ex ante hypothetical bias mitigation methods

Method	Seminal studies	Field	Application	Description
Cheap talk	Cummings & Taylor (1999)	Environmental economics	CV	Explains the problem of hypothetical bias to study participants prior to administration of a hypothetical question.
Consequentiality	Carson & Groves (2007)	Environmental economics	CV	Informs the study participants about the possibility that results of the study may affect an outcome they care about.
Opt-out reminder	Loomis et al. (1994)	Food choice	CV, DCE	Reminds study participants to choose the opt-out alternative if the prices of the experimentally designed alternatives are greater than what one would actually pay in actual shopping situations.
Solemn oath	Jacquemet et al. (2013)	Environmental economics	CV	Asks participants to promise that they will reveal their true preferences when responding to the choice questions.
Honesty priming	De-Magistris et al. (2013)	Food choice	DCE	This task induces respondents to reveal their true preferences as per the effect of a mental activation process.
Inferred method	Lusk & Norwood (2009)	Food choice	CV	Asks respondents to make their choices based on what they think an average consumer would choose.

Abbreviations: CV, contingent valuation; DCE, discrete choice experiment.

A second strand of the literature compares multiple correction methods either within purely hypothetical experimental settings or by comparing real and hypothetical DCEs. These studies also draw divergent conclusions. De-Magistris & Pascucci (2014) compared the performance of using oaths or cheap talk scripts using a hypothetical DCE on insect-based food. They found that using a solemn oath reduced hypothetical bias more than cheap talk scripts. In a Spanish study on almond choice, De-Magistris et al. (2013) compared honesty priming and cheap talk using both hypothetical and real DCEs. They report that honesty priming outperforms cheap talk scripts in reducing hypothetical bias, as it produced lower WTP estimates in hypothetical settings compared to real experiments. These findings are confirmed in a purely hypothetical DCE study on tomato selection in Nigeria by Bello & Abdulai (2016). Together, the evidence from these studies suggests that honesty priming is more effective than cheap talk scripts at mitigating hypothetical bias. But cultural and experimental settings might affect results. For example, quite different conclusions are drawn from research by Lin et al. (2019) who conducted a DCE in China. In an online survey to assess Chinese consumers' WTP for pork loin, they compared purely hypothetical choices with and without three correction methods (cheap talk, honesty priming, and oath). Results from this study showed no significant difference in WTP estimates between the hypothetical stated choices with and without the three correction methods.

Overall, the results from these two strands of literature suggest that the effectiveness of ex ante mitigation techniques is both product and context dependent. These results further highlight other research gaps. For instance, currently there are no studies comparing all existing ex ante mitigation techniques. Meta-analysis studies allow researchers to combine results from prior studies and thus compare the performance and efficacy of multiple ex ante correction methods.

Indeed, a recent meta-analysis study by Penn & Hu (2018) suggests that combining cheap talk, consequentiality, and certainty follow-up can significantly contribute to mitigating the magnitude of hypothetical bias in DCEs. Yet, differences in geographical and cultural contexts, research designs, and participant pools can often limit the comparability of study results and thus affect the external validity of findings generated by meta-analysis studies. Perhaps more importantly, most of these prior applications have been conducted in developed or western countries. Thus, very little is currently known about which correction methods reduce or eliminate hypothetical bias in the operational conditions prevailing in DCEs conducted in developing countries (Scarpa et al. 2013, Bello & Abdulai 2016, Alemu & Olsen 2018). This is despite well-known issues related to the quality and availability of secondary data and an increasing use of DCEs in developing countries. The investigation of hypothetical bias in developing countries is an area that deserves further investigation.

2.2. Real Food Choice Experiments

To avoid hypothetical bias, many researchers use real (or non-hypothetical) discrete choice experiments (R-DCEs) (Lusk & Schroeder 2004; Alfnes et al. 2006; Moser et al. 2014; Bazzani et al. 2017a,b). Participants in R-DCEs are presented with repeated choice tasks, each composed of experimentally designed alternatives differing by attributes and their levels. An opt-out alternative is also typically included. However, unlike in DCEs, in R-DCEs subjects are provided with actual (non-hypothetical) economic incentives, involving real products and real money. At the end of the choice exercise, one of the choice tasks is randomly selected as a binding choice for payment, and the respondent is asked to purchase the chosen product unless the opt-out option is selected. It is expected that the use of actual products and real market incentives, although somewhat diluted by the randomness factor, provides a sufficient incentive to induce subjects to truthfully reveal their preferences (Lusk & Schroeder 2004, Alfnes et al. 2006, Chang et al. 2009).

Empirical applications include food preference analyses as well as testing of implications from economic theory. Examples of R-DCE studies include those by Alfnes et al. (2006), who assessed consumer's WTP for salmon with various degrees of flesh redness across different information settings, and Bazzani et al. (2017b), who tested the commitment cost theory (Zhao & Kling 2004) in static and dynamic choice settings and its effects in food purchasing decisions. However, R-DCE applications are necessarily limited to private goods and existing food products. This is a considerable limitation because many food DCEs are implemented to evaluate novel food products and/or food labels not yet available in actual markets as well as new food technologies not yet fully developed or market ready. More recently, Fang et al. (2021) explored whether the use of novel technologies such as virtual reality (VR) could reduce or mitigate hypothetical bias and thus improve the validity of findings from DCEs. Results from this study find that the use of VR in DCEs does not eliminate hypothetical bias, but it can reduce it significantly, compared to the more common computer-based delivery of choice questions via pictures and plain text. Yet, this evidence only holds for participants who did not exhibit high VR discomfort. This evidence aligns well with similar studies in environmental economics based on VR, enhanced reality (Bateman et al. 2009), and virtual environments (Matthews et al. 2017).

3. FOOD CHOICE DESIGNS AND INTERNAL VALIDITY: CURRENT PRACTICES AND ISSUES

Devising an experimental design for a DCE that can accurately identify and efficiently estimate all desired effects requires a sequence of many decisions, which are made at different stages during the development of the DCE survey instrument (Louviere et al. 2000, Hensher et al. 2015). This section reviews relevant methodological contributions within the food choice domain and discusses a few issues that we believe can help improve the way food DCEs are designed and implemented.

3.1. Elements of the Experimental Design and Design Dimensionality

The selection of alternatives, attributes, and attribute levels represents the first step researchers undertake when designing a food DCE study. This step is driven by the research objectives, and it is typically informed by prior literature, pilot studies, and/or qualitative research as, for example, the opinions collected during well-organized and conducted focus groups.

Other decisions concern design dimensions, such as the number of choice tasks in the sequence, the number of alternatives, the number of attributes, the number of attribute levels, and the range of attribute levels to include in the study. Design dimensions determine choice task complexity and, when this is high, they may pose a cognitive burden on subjects. Complexity may influence the way participants respond to the preference-elicitation questions (Swait & Adamowicz 2001, DeShazo & Fermo 2002), inducing them to engage in undesirable practices, such as making choices by employing various effort-saving heuristics (Hensher 2006). Indeed, evidence from the field of transport and environmental economics suggests that different design complexity may affect welfare estimates and other data quality aspects (Hensher et al. 2001, DeShazo & Fermo 2002, Arentze et al. 2003, Caussade et al. 2005, Rose et al. 2009, Oehlmann et al. 2017, Campbell et al. 2018).

Understanding the extent to which task complexity affects welfare estimates under standard econometric assumption (DeShazo & Fermo 2002) is an important methodological issue also in food DCE applications, as it has significant implications in terms of survey designs and reliability of estimates. When WTP estimates are sensitive to changes in design dimensionality (Meyerhoff et al. 2015), researchers should be wary about whether their DCE design provides the optimal level of complexity. Excess complexity may affect robustness and reliability of findings (Lusk 2003) and induce tiredness in respondents. The relationship between measures of choice task complexity and latency, which is the time respondents take to complete choice tasks, is often taken as evidence of internal validity (Atzori et al. 2022). The rationale is that as complexity increases, response latency should also increase, everything else equal. These findings provide a criterion for optimally sequencing choice tasks at the respondent level, suggesting that tasks should be presented in order from the least to the most complex, using ex ante measures of complexity. The following subsections outline common procedures used by researchers to design food choice studies. We also discuss recent methodological contributions on survey designs and implementation.

3.1.1. Number of attributes. Only two food choice studies have explicitly addressed the effects of varying the number of attributes in the food choice literature: those by Gao & Schroeder (2009) and Caputo et al. (2017). Both studies focus on the effects of varying the number of attribute information on the stability of WTP estimates when both cue and independent attributes[6] are present on steak products. Gao & Schroeder (2009) implemented between- and within-subject designs. The between-subject design consisted of four DCE surveys: two with the presence of independent

[6]Caputo et al. (2017) define cue and independent food attributes as follows: A cue attribute can be described as one that embeds quality features of other attribute information that is not explicitly detailed in the product's description. For example, country of origin is considered a cue attribute, as it can be seen by consumer as a proxy of other quality features such as freshness, safety, etc. In contrast, an independent food attribute refers to physical aspects of a food product and this is not perceived to embed further quality cues by consumers.

and cue attributes and two without the cue attribute.[7] In each of these surveys, respondents faced two sequences of choice questions (within-subject design): The first sequence of choice questions was designed with three or four attributes and the corresponding second sequence of choice questions with four or five attributes. The authors found that (*a*) the marginal WTP estimates for a given attribute are sensitive to the number of attributes used to characterize the composite good of interest and (*b*) that marginal WTP estimates for cue attributes are more sensitive to changes in the number of attributes than independent attributes.

Similar results were documented by Caputo et al. (2017). The authors implemented the same designs and between- and within-subject survey approaches. However, unlike Gao & Schroeder (2009), they only focused on the effect that adding independent food attributes had on the robustness of the marginal WTP estimates of a cue attribute. They conducted three DCEs, with the number of attributes increasing from three to four, four to five, and five to six, respectively. Results from this study, obtained with a sequence-consistent panel estimator, indicate that changes in marginal WTP estimates depend not only on the number of attributes but also on the functional role played by the cue and independent attributes. The authors argue that whether the attribute is perceived by consumers as a quality cue or independently of other quality features is a function of the design dimension. In a design with a smaller number of attributes (from 3 to 4 and from 4 to 5), changes in marginal WTP estimates are associated with both cue and independent attributes, but when the dimension of the attribute space is higher (from 5 to 6), only the cue attribute remains affected.

Taken together, the results from these two studies show that the way consumers process attribute information depends not only on their number of attributes included in the design but also on the attribute type (e.g., cue versus independent) and their functional roles (e.g., how consumers perceive the experimentally designed food attributes). More specifically, evidence to date indicates that the use of small designs (fewer than 5 attributes) can lead to bias estimates (overestimation) of both the cue and independent attributes, if included in the study. Although it is not yet addressed in the food choice literature, the order of appearance of extreme attribute levels in the sequence of choice tasks can also influence choice behavior in public goods provision (Day et al. 2012). Therefore, future research could look at the internal dynamic of the sequence of choices to derive guidance on the adequate sequence of attribute (and level) appearance.

3.1.2. Number of attribute levels and value range. In addition to selecting a suitable number of attributes for the experiment, it is also important to choose appropriate attribute levels, which accurately reflect both the context in which the experiment is conducted and the type of inference the derived choice probability model is meant to inform.

Selection of the number of attribute levels and their value range are critical aspects for the price attribute, which is generally incorporated in food DCEs as one of the relevant attributes to translate marginal utility estimates from probabilistic discrete choice models into money metric WTP estimates. Product price levels are typically selected to encompass the range of actual prices in grocery stores. These can be obtained from official sources (e.g., the US Department of Agriculture/Economic Research Service in the United States, EUROSTAT in the European Union, and similar national statistical offices) through an exploratory analysis whereby actual prices are collected in different retail stores (e.g., grocery stores, gas stations, discounts) and/or by conducting focus groups with consumers. The latter is particularly important when considering DCEs on

[7]In addition to the price, the authors selected guaranteed tender, guaranteed lean, days before sell-by date, and enhanced omega-3 fatty acids as independent attributes, and certified US product as the cue attribute.

novel food products still unavailable in the market and when the focus is on WTP. Yet, research on the optimal number of price levels and the width covered by the price range is scarce, particularly within the food choice literature. Most food choice studies include three to five price levels (Chang et al. 2009; Lusk & Schroeder 2004; Scarpa et al. 2013; Caputo et al. 2013a,b; Caputo 2020), although a few studies have also conducted food DCEs with only two price levels (Ahn & Lusk 2021) or more than five price levels, especially in DCEs on wine selection (Tait et al. 2019).

Only the study by Mørkbak et al. (2010) examined the effects of varying price levels on marginal WTP for food quality attributes. The authors employed a between-subject approach to investigate whether increasing the highest price level by 50% in the treated group produces different marginal WTP estimates for quality attributes in minced pork. Results show that such an increase leads to higher WTP estimates (up to 68%) and are in line with results from studies in other fields such as environmental economics. Carlsson & Martinsson (2008) and Kragt (2013) found that an upward shift of the price levels resulted in higher WTP estimates in their estimates of consumer WTP to reduce unplanned power outages and for natural resource management, respectively. Contrasting findings show no differences between estimates of population means of marginal WTP across different price vectors that are documented in energy economics studies (see, e.g., Aravena et al. 2014, Glenk et al. 2019). Price vector effects in food choice might benefit from further research.

3.1.3. Number and types of alternatives and choice tasks. No prior food DCE has examined the effects of design complexity produced by a different number of choice tasks and/or alternatives on choice behavior. Most food choice experiment studies have asked respondents to evaluate between six and twenty-five choice tasks, with eight and twelve tasks being the most frequent. From a statistical perspective, more tasks are desirable, but from the perspective of the cognitive burden on respondents the opposite is true.

The number of choice tasks researchers include in DCE designs raises two important issues. The first is a model identification issue, as the number of choice tasks determines the degrees of freedom[8] required in estimation; a higher number of choice tasks (i.e., a longer panel) implies higher degrees of freedom in the estimation of models. Longer panels are useful especially when the goal is forecasting individual choice probabilities by each respondent or when individual marginal WTP estimates of attributes are of interest (Bech et al. 2011, Sarrias & Daziano 2018). The second issue regards the potential for poor data quality. The length of the choice exercise may affect the scale of the Gumbel error, inducing heteroskedasticity. The error scale goes quickly up (variance goes down) after the first choices and then tends to plateau until respondents' fatigue kicks in and brings it down again (Scarpa et al. 2011). So, designs with too many choice tasks may affect the completion rate, induce random choices due to fatigue, and/or lead respondents to deal with the volume of information they are presented with by employing a coping strategy. For example, designs may induce them to ignore some information, which moves away from the desired behavior of fully compensatory trade-offs. Tired respondents may act as so-called cognitive misers and use various decision rules to reduce the cost of cognitive effort (Simon 1955). Results from choice modeling studies outside the food choice literature indicate that subjects can adequately handle more than 16 choice tasks with minimal or no impact on response rates, fatigue, detachment, and welfare estimates (Hensher et al. 2001, Louviere 2004, Carlsson & Martinsson 2008, Bech et al. 2011).

[8]Hensher et al. (2015) define the degree of freedom as $S \times (J-1)$, where S is the number of choice tasks and J the number of alternatives in each choice task.

As for the number of alternatives one should use in a DCE context, a similarly careful answer must be given. The number of alternatives is highly dependent on a variety of factors, most prominently whether the design alternatives are labeled (e.g., immediately recognizable from the rest because of distinctive features such as their brand or color) or unlabeled (e.g., distinguishable exclusively on the basis of the attribute levels).[9] Labeled designs typically call for a larger number of alternatives than unlabeled ones. Most unlabeled choice tasks (or choice questions) offer the respondent only two alternatives to choose from, while labeled choice experiments typically use between four and nine different alternatives (Lusk 2017; Caputo et al. 2018a, 2020; Van Loo et al. 2020) that respondents can choose from. The inclusion of a none or no-purchase alternative is the norm in both labeled and unlabeled food choice studies to reflect that some respondents might choose this option when shopping (Lusk & Schroeder 2004). It is unclear, but certainly worth investigating, whether the source of biases discussed in other fields (Oehlmann et al. 2017, Carson et al. 2020) and related to the presence of opt-outs, such as no-choice or status-quo, in the choice task maintain their relevance in food DCEs. Evidence from induced-value experiments in the environmental economics literature indicates that choice questions with three and two alternatives (*a*) provide convergent estimates of WTP to induced values, and (*b*) nonstatus quo alternatives are associated with lower WTP estimates. This is in keeping with the experimental results and the suggested mechanism associated with designs including status-quo/no-buy alternatives in all choice tasks proposed by Carson et al. (2020). This theme is in need of further investigation in food choice studies.

3.2. Experimental Designs

To generate choice tasks (or choice questions) to use during experiments, the DCE literature proposes various experimental designs including only a subset (or fraction) of choice tasks from the full factorial design. These designs differ depending on how attributes and attribute levels are combined across alternatives and subsets of repeated choice tasks and whether designs are generated to incorporate only attribute main effects or also their interaction effects (see ChoiceMetrics 2021). The fractional factorial orthogonal design is the most used design in food choice studies. This design generates uncorrelated and balanced attribute levels. Limitations of this design include (*a*) a likely loss of orthogonality in the final sample, due to incomplete responses across the sequence of choice questions, (*b*) presence of dominant alternatives that imply no trade-offs, and (*c*) existence of duplicates. Both (*b*) and (*c*) limit the amount of preference information gained from the observed choices.[10] The study by Lusk & Norwood (2005) is the only food choice study investigating the impact of random and fractional orthogonal designs on welfare estimates from the multinomial logit (MNL) model. Using a Monte Carlo simulation, the authors documented three key findings: (*a*) Experimental design can influence the accuracy

[9]In a labeled design the purchase alternatives represent specific products (i.e., ketchup, mustard, mayonnaise, and ranch dressing) or brands of the same product (for ketchup: Del Monte, Red Gold, Heinz, and Hunt's). These product/brand alternatives do not change across choice questions. In each choice question respondents are asked to evaluate the product/brand alternatives at the posted price and the presence/absence of the other attribute levels included in the experimental design. In an unlabeled design, the purchase alternatives are generic; i.e., they do not refer to any specific product or brand. In each choice question, respondents are asked to evaluate different purchase alternatives, which are characterized by the presence/absence of the experimentally designed attribute levels. Therefore, they are typically labeled as Product A and Product B or Option A and Option B.

[10]Researchers could discard replications and/or dominant choice tasks. Yet, incomplete fractional orthogonal designs are found to underperform compared with other designs (Walker et al. 2018).

of WTP estimates, although none of the studied designs produce biased WTP estimates; (*b*) designs incorporating attribute interactions generate more precise WTP estimates than designs without interactions, especially in the case of a random design;[11] and (*c*) large sample sizes increase efficiency and hence can mitigate issues arising from poor experimental designs.

More recent food choice applications proposed alternative methods to extract practical fractions from the full factorial. These include experimental designs that are D-optimal and efficient designs. The D-optimal design is constructed to maximize the differences in attribute levels across alternatives. This is achieved using algorithms called design generators (see Street et al. 2005). This approach implies that the generated design is free of duplicates and respondents are forced to make trade-offs (this may lead to lexicographic choice behavior). From a statistical perspective, D-optimal designs are limited orthogonal designs in that they are orthogonal within an alternative but have (often perfect negative) correlations across levels in the alternatives.

Of great interest are design generators specific for increasing the efficiency of the MNL, which is the workhorse model used in data analysis from DCEs. MNL D-efficient designs minimize the determinant of the variance-covariance matrix of the model parameter estimates (or functions thereof) within the set of the design choice tasks. The computation of such determinants requires some assumptions about the values of the utility parameters, the values of which are the final goal of estimation. These designs can be generated using two different approaches. The first approach assumes that the parameter priors are nonzero and known with exact certainty by the researchers (e.g., Bliemer & Rose 2005, Scarpa & Rose 2008). Therefore, these designs are truly efficient only in ex ante sample data collection, because after the collection of the data the estimates of the utility parameters may significantly depart from the assumed values (or priors). Hence, wrongly specified priors typically lead to some degree of inefficiency ex post (i.e., after the sample is collected; see Walker et al. 2018). The second approach relaxes the assumption of perfect a priori knowledge of the parameter priors and assumes, using a Bayesian approach, that prior parameter values are distributed ex ante according to some parametric distribution to be specified by the researcher (Sándor & Wedel 2001). This approach attenuates the potential bias due to incorrect a priori assumptions on parameter values by formally addressing ex ante uncertainty. Findings from previous studies have shown that Bayesian efficient designs derived under mildly correct a priori assumptions are able to produce greater precision in functions of parameter estimates, such as marginal WTPs with smaller standard errors even at smallish sample sizes (Scarpa et al. 2007). However, evidence from recent studies reminded practitioners that wrongly specified priors may lead to a significant loss in efficiency also in Bayesian D-efficient designs (Walker et al. 2018).

Although precision of parameter estimates is a good design criterion, sometimes other features are of interest to researchers. Efficient designs can also be derived for specific parameter functions, such as marginal rates of substitutions (e.g., WTP for attributes) as shown by Scarpa & Rose (2008), and Bayesian efficient designs can be specialized to marginal WTP estimates, as shown by Vermeulen et al. (2011). D-efficient designs minimize the determinant of the variance covariance matrix of the parameter estimates for the MNL model and are often used in choice experiments because they have recently been found to also be efficient in semiparametric estimation (Scarpa et al. 2021b). Similarly, for accurate parameter estimates the A-optimality criterion can be used, which minimizes the average of the parameter variances and, unlike D-efficiency, A-optimality ignores the covariance factors. Other criteria, such as G- and V-optimality, focus on accuracy of predictions, rather than utility parameters or marginal WTPs (Kessels et al. 2006). While G-optimality focuses on minimization of the maximum of prediction variance, V-optimality

[11]The random design consists of randomly drawn choice tasks from a full factorial design.

focuses on minimization of the average of prediction variance. All of these criteria can be given a Bayesian form, once assumptions on prior distributions are feasible. Despite their evident advantages, none of the above criteria are of common use in food DCEs. Although efficient algorithms exist for their derivation (Kessels et al. 2009), these are complex and time consuming to execute.

3.3. Sample Size Requirements and Statistical Power

Sampling theory for maximum likelihood estimates can be used to define sample size requirements in terms of statistical significance of the estimates produced from MNL models. Bliemer & Rose (2005) proposed a number of measures that can be used to evaluate a priori the statistical efficiency of the parameter estimates an experimental design generates, while Scarpa & Rose (2008) provided guidance on how to evaluate the performance of a design after sample collection. Among others, the *S*-error is a statistical measure that provides the theoretical minimum sample size necessary to obtain asymptotically significant parameter estimates (assuming a given confidence level or *t*-value and a set of parameter values) of the MNL model. These are a priori sample size indications and specifically derived from data collected from a given experimental design and set of parameter assumptions (see Bliemer & Rose 2005). The **Supplemental Appendix** provides an illustrative example on how to obtain the *S*-statistics in Ngene, software for designing choice experiments.

Supplemental Material >

Of course, in the context of food choice, researchers can be interested in obtaining indications about adequate sample sizes for identifying treatment effects above given thresholds (effect size for the power of the identification test) for marginal WTPs, probabilities of selection, or marginal effects. Given model specifications and associated variances for the estimates at hand, the concept of the *S*-statistic can be adapted to all these cases. When moving away from the standard MNL model specifications and venturing, for example, in mixed logit (MXL) models of various types, efficient experimental designs and power tests for effect identification become considerably more complicated owing to the additional sources of estimation variance. For example, apart from sampling variance, other sources of variation include simulation draws, alternative distributional assumptions of random parameters, different correlation structures between random parameters, and the choice of convergence algorithm, iteration stopping rules, and order of variables (Palma et al. 2020). However, guidance on sample size requirements can still be obtained from adequately designed Monte Carlo experiments. Indeed, further work is required to establish standard practices to establish sample size requirements and statistical power in DCEs in general and food DCEs in particular.

4. MODELING

A steady stream of research from the fields of transportation, environmental, and health economics has examined different discrete choice models and utility specifications to provide more reliable, behaviorally credible, and reasonable estimates of WTP. Discrete choice models with random coefficients for panel data, such as the repeated choice MXL model (Revelt & Train 1998; Train 1998, 2009) and its expansion, the MXL model with an error component for the no-choice/no-buy alternative model (MXL-EC) (Scarpa et al. 2005, 2007), now represent the state of the art in the broader field of choice modeling and food choice (Lusk & Schroeder 2004; Alfnes et al. 2006; Caputo et al. 2013a; Scarpa et al. 2013) and are also available for Bayesian estimation (Balcombe et al. 2009, 2014, 2015). Researchers utilizing MXL models face two main issues: whether models should be specified in preference space or WTP space and whether parametric or more flexible nonparametric distributions should be used for random taste coefficients. The following subsection discusses these two issues in more detail.

4.1. Preference Space Models and Parametric Distributions

Most food choice studies utilizing MXL models are estimated using utility functions specified in preference space and assume parametric distributions for random parameters.[12] The coefficients of nonprice food quality attributes are usually treated as random and assumed to follow a normal distribution to accommodate positive or negative values of preferences for food product attributes. The task then becomes that of estimating a mean vector and a variance-covariance matrix. The price coefficient, in contrast, is often assumed to be fixed across individuals. After estimation, these coefficients are used to derive marginal WTP estimate for each attribute, which is obtained as the negative ratio of the nonprice food attribute coefficient and the price coefficient (Train & Weeks 2005). The assumption of a fixed price coefficient facilitates estimation and allows the distribution of WTP values to be the same as the one for nonprice attribute coefficients (normal distribution). However, a constant price coefficient presents two main limitations: (*a*) It implies a constant marginal utility of money across individuals, which runs counter to economic intuition given that the same monetary unit can have different values for individuals with different endowments and budget constraints, household composition, and stage in life; and (*b*) it implies a constant standard deviation of unobserved utility (i.e., the scale parameter) over individuals (Scarpa et al. 2008).

To overcome these limitations within a preference space framework, two different approaches have been adopted in the food choice modeling literature so far. The first pertains to the specifications of utilities in preference space where the price coefficient is also assumed to vary across individuals in an unobserved manner, according to same distribution. The second approach follows that of Morey et al. (2003), who use group-specific price coefficients, where the group can be defined as belonging to a specific income class, household size, etc.

When treating the price coefficient as random in a preference space framework (the first approach), researchers need to select a distribution for the price coefficient and estimate the parameters of such distribution. In food DCE studies, the price coefficient is generally expected to be negative because higher prices are generally perceived as bad, which is accompanied by a comparatively lower likelihood of purchase.[13] Therefore, bounded (below zero) distributions, such as the lognormal or the constrained triangular, might be more desirable, as they constrain the parameter sign to be positive. To achieve the required negative value, the price variable can simply be specified as negative. However, the presence of fat tails in lognormal distributions and a positive density close to zero do not elude the problem of obtaining implausible distributions of marginal WTPs when computing the ratio between the attribute coefficient and the price coefficient. This is the so-called exploding implicit price problem (Giergiczny et al. 2012). As such, the assumption of a price (or cost) coefficient following a constrained triangular distribution (Hensher & Greene 2003) is also common in food choice studies (Alfnes et al. 2006, Scarpa et al. 2013, Van Loo et al. 2020). Such a distribution limits the range of variation and ensures that the sign of the price estimate is behaviorally plausible throughout the entire distribution (Hensher et al. 2015). When

[12] The distributions of the food attribute coefficients or marginal utilities are assumed ex ante by the researchers according to their expectations.

[13] Conversely, it is also possible that some individuals may view product price as a proxy for food quality and therefore value high prices more than low prices. If such a "mixed price effect" is expected, then a normal distribution (or truncated normal, -1.96 to 1.96) could be empirically appropriate to use when modeling the price coefficient in food choice models; a normal distribution is unbounded, implying that some share of the population actually prefers higher prices. However, it is well known in the choice modeling literature that because the normal distribution assigns positive density around zero, this can preclude the calculation of finite moments for the marginal WTP distribution.

estimating choice models with a price coefficient assumed to be triangularly distributed, the spread can be constrained to equal the mean. Hence, only a single parameter is estimated that represents both the mean and the spread. Therefore, one limitation of such distributions is that they prevent the analysis of correlation patterns across attributes and subsequently also preclude testing the statistical significance of the scale parameters because the location parameter is constrained to be equal to its scale (Campbell et al. 2009). Day et al. (2012) present a theorem that allows researchers to test whether the derived WTP distributions have finite moments when assuming random coefficient for the price/cost attribute.[14] They show that (*a*) the normal distribution implies infinite moments for the distribution of respondent's WTP; (*b*) in the truncated normal and constrained triangular distributions, the existence of the inverse moments depends on the actual estimated parameters of the distribution; and (*c*) the lognormal distribution has all inverse moments in their basic specifications (i.e., with the domains between 0 and infinity for the lognormal). A recent proposal to use the (negative) shifted lognormal as a distributional assumption for random price in models with preference-space utility looks promising (Crastes dit Sourd 2021).

4.2. WTP-Space Models

Daly et al. (2012) suggest that models specified in the WTP space represent the most straightforward solution to ensure distributions of WTP values that have finite moments, as this avoids the need to consider the distribution of an inverse coefficient. In addition, the utility coefficients of choice models in the WTP space can be directly interpreted as marginal WTP estimates. Results from a number of studies in the fields of transportation and environmental economics suggest that the use of WTP-space estimation (*a*) is more practical for derivations of welfare estimates and for accounting for interpersonal scale variation (Scarpa et al. 2008); (*b*) provides more reasonable distributions of WTP (Train & Weeks 2005) and stable WTP estimates (Balcombe et al. 2009); (*c*) may result in a significant improvement in model fit (Scarpa et al. 2008); and (*d*) may be used to test restrictions on the dispersion of random parameters directly during estimation using standard MXL tests (Thiene & Scarpa 2009), such as the Wald, score, and likelihood ratio tests. Other studies, however, have shown that models in the WTP-space underperform in terms of fit to the data relative to those with utility in the conventional preference space estimation (Train & Weeks 2005). According to Rose & Masiero (2010), differences across these studies might be due to the nature of the data set used or to the experimental contexts. Differences might also be due to the various assumptions underlying the distribution of the price coefficient. As pointed out by Train & Weeks (2005), research is needed to identify distributions that fit the data better when applied in the WTP space and/or provide more reasonable distributions of WTP when applied in the preference space. Which specification works better remains a data set–specific empirical question at this stage. All of these models require sufficiently large sample sizes and adequate designs to capture preference heterogeneity and potentially test alternative distributional assumptions for random taste parameters.

4.3. Semiparametric Distributions in the Preference and WTP Spaces

Recent developments in the choice modeling literature include the logit-mixed logit (LML) model developed by Train (2016). The model allows for semiparametric mixing distributions, thus accounting for asymmetry and multimodality, and it is computationally much easier to implement

[14]They showed that if the distribution of the price/cost parameter has a positive density at zero, then the resulting ratio distribution for WTP does not have finite moments.

than previous semiparametric models with the same capacity (Scarpa et al. 2008). The LML model has been used in food choice studies by Bazzani et al. (2018), Caputo et al. (2018b), Kilders & Caputo (2021), and Scarpa et al. (2021b). Collectively these studies suggest that (*a*) preferences for food quality attributes are unlikely to be normally distributed. Indeed, they can often display bimodality and asymmetry (Caputo et al. 2018b), especially in the case of preferences for new food technologies (Kilders & Caputo 2021). (*b*) Flexible distributions can provide more reliable and insightful estimates of WTP distributions (Bazzani et al. 2018). Scarpa et al. (2021b) study the trade-off between bias and variance in the choice between the MXL model with random normal parameters and a semiparametric LML model. They show that, under frequent operational conditions in terms of attributes and levels, the sample size at which the trade-off moves in favor of the LML is likely to be more than 500 respondents. At smaller sample sizes the effect of the variance of the estimator dominates the reduction in bias, and one is likely to be better off using the MXL. Although all studies using LML models in food choice studies have used the WTP-space utility, there is no reason to believe that similar conclusions should not extend to preference-space utility specifications, which have been explored by Bansal et al. (2018).

5. BEHAVIORAL INSIGHTS

This section discusses other modeling approaches that food economists have employed to incorporate behavioral insights into discrete choice models. Given existing high-quality reviews of neuro-economics research (see, e.g., Palma 2022 for a review on neurophysiological measures that could be used in consumer food choice and Van Loo et al. 2018a for a review on the relationship between visual attention and food choice), here we focus on behavioral factors that stem from the psychology and behavioral economics literature.

5.1. Behavioral Factors Influencing Decision-Making Processes

Modeling food choices requires an adequate understanding of the processes underlying human decision making. For instance, an implicit assumption in DCEs is that individuals make choices only based on prices and food product attributes included in the experiment. However, individual choices can be influenced by other factors that even well-designed experiments cannot easily control for. Caputo et al. (2018a) show that prior knowledge and expectations by subjects, such as subjective reference points (e.g., benchmark prices) and external pricing information, can significantly affect choice. This evidence has significant implications for the design and practice of DCEs, as they show that experiments are not conducted in a vacuum. In addition, given fluctuations in market prices, when shopping for groceries, consumers often face uncertainty over whether they are indeed getting a good deal. Caputo et al. (2020) move from this concept and develop a novel model that incorporates reference price uncertainty into food choice. The authors show that ignoring reference price uncertainty in food demand analysis leads to biased market share estimates. They offer a new empirical framework to model uncertainty in food choice studies. Collectively, results from these studies indicate that food DCE applications should consider reference points as an explicit aspect of experimental designs to prevent pricing bias and account for reference price uncertainty when estimation of market shares is the focus of the study. Both aspects should be taken into consideration and further evaluated in future research. Along the same vein, but with a slightly different perspective on the sources of uncertainty, Scarpa et al. (2021a) adopt the approach suggested by Manski (1999, 2004) and operationalized by Blass et al. (2010) in the context of choosing extra virgin olive oils. They explore the notion of resolvable uncertainty, which is due to differences in information sets faced by respondents between two types of choice: stated and real. They focus on inferring the information gap between knowledge available to subjects at the time when choice statements are collected and when making a real purchase, making use of statements on probabilities of choice, rather than preferred choice.

In addition, how research subjects process food product information during DCEs depends on the complexity of the experimental design. Although complex designs that resemble actual shopping situations can increase realism and hence be tempting to researchers, they also challenge respondents and complicate both survey operations and statistical estimation. Evidence from both psychology and behavioral economics indicates that consumers may use heuristics (Simon 1955, Payne et al. 1992), a mental shortcut that allows people to make quick and effortless decisions. Indeed, several DCEs have shown that consumers tend to ignore (not attend to) some of the attribute information presented to them during the choice exercises (Gilbride & Allenby 2004, Scarpa et al. 2010, Hensher et al. 2012). This behavior, known in the choice modeling literature as attribute-nonattendance (ANA) behavior, can lead to inadequate model specification and thus produce poorer statistical fit and biased WTP estimates (Campbell et al. 2008, Scarpa et al. 2010).

Methodological contributions in this research domain include various methods developed by researchers from different fields to model and infer information on ANA. Some studies propose the use of self-reported ANA statements that can be collected at the end of the choice exercise (stated serial ANA) or at the end of each choice tasks (stated task ANA). Other studies, following Scarpa et al. (2009), suggest inferring ANA behavior using suitable latent class models based on panel data (Caputo et al. 2013a). A much-debated issue in food choice research and elsewhere concerns the validity of stated serial and choice task ANA elicitation methods, as well as the concordance between stated and inferred ANA and their relationship with design complexity. Food choice applications include those by Scarpa et al. (2013), Balcombe et al. (2015), Caputo et al. (2018c), and Van Loo et al. (2018b). Scarpa et al. (2013) explored the validity of stated serial ANA on beef and chicken selection and its concordance with inferred ANA. Results from this study show that that the constrained latent class panel models outperform MXL models with random parameters in terms of fit to the data and correct downward marginal WTP estimates for food attributes. Caputo et al. (2018c) validated self-reported ANA statements for serial choice tasks and explored the concordance of stated methods with that of the inferred ANA. Results show that self-reported measures on choice task ANA are more congruent with model estimates than those for serial ANA, as well as with inferred ANA. In a DCE on potato preferences, Kaye-Blake et al. (2009) developed an elegant computer-based survey in which respondents were given choice tasks with concealed attribute levels that could be disclosed on demand by respondents if they were found of interest. By doing so, respondents disclosed the structure of their individual utility showing to the researcher what attributes mattered at each choice task. Their work is a precursor of eye-tracking methods and demonstrates that respondent-specific utility structures can be retrieved with careful data collection. Balcombe et al. (2015) and Van Loo et al. (2018b) used various eye-tracking measures (e.g., fixation count cutoffs) to account for visual ANA. Balcombe et al. (2015) found that although informative, stated ANA statements diverge from visual ANA statements. Van Loo et al. (2018b) found that visual ANA is attribute dependent and concluded that using eye tracking does not always provide an adequate measure of ANA. More recently the inferred ANA approach was used to measure inattention bias in food DCEs. For instance, Malone & Lusk (2018a) estimated a latent class logit model with all parameters in one class restricted to zero. The authors then used the estimated share of observations falling in the class with null parameters (representing totally random choices) as a diagnostic measure of total inattention to food attribute levels in the choice task alternatives.

Other behavioral factors have been found to influence food decision making. These include personal values (Grebitus et al. 2013b), personality traits (Grebitus et al. 2013a, Bazzani et al. 2017a), time preferences (De Marchi et al. 2016), beliefs/perceptions (Lusk et al. 2014, Malone &

Lusk 2018b), and ethnocentrism (Van Loo et al. 2019), among others. For example, De Marchi et al. (2016) demonstrate that consumer demand for healthy and environmentally friendly food products is driven by time preferences, that is, how consumers weigh the value of future events. These authors discuss the importance of developing food policies and programs that can educate people about the long-term benefits of healthier and more sustainable foods. Grebitus et al. (2013a) show that consumers' personalities influence choice behavior differently in real and hypothetical food choice environments, suggesting that personality could explain a significant portion of hypothetical bias. Van Loo et al. (2019) find that high consumer ethnocentrism leads to higher preferences for domestic food, as also evidenced by early food choice studies documenting home-biased and patriotic preferences (Scarpa & Del Giudice 2004, Loureiro & Umberger 2007, Menapace et al. 2011). Taken together, these results clearly show that researchers need to be wary of potential behavioral influences on their DCE and, if possible, take appropriate steps to account for those influences within their experimental design and/or modeling approach.

5.2. Endogeneity Bias in Choice Experiments

Most research investigating whether and how behavioral factors influence food choice behavior used self-reported auxiliary data collected during the experiment in the form of Likert or semantic scales, probability statements, or multiple-choice questions. For example, Grebitus et al. (2013a) employed a validated scale to determine consumer personality traits, Lusk et al. (2014) developed probability statements to elicit consumers' subjective beliefs, and Scarpa et al. (2013) and Caputo et al. (2018c) utilized Likert scales and multiple-choice questions to collect and explain self-reported ANA statements.

The incorporation of self-reported auxiliary data as "error-free exploratory predictors of choices" (Hensher et al. 2015, p. 929) into random utility models can raise endogeneity problems. In random utility models, endogeneity arises if variables that enter systematic utility components are correlated with random utility components due to measurement errors and/or the built-in causal effect between food choices and latent variables, such as beliefs/perceptions/self-reported ANA. However, the definition of endogeneity is complicated when choice data are generated through an experimental design, and its identification can be problematic due to the challenge of finding plausible instrumental variables that allow researchers to distinguish behavioral drivers from preferences. Recent food choice studies have employed various empirical strategies to incorporate behavioral auxiliary data into choice models while also correcting for potential endogeneity problems. Some food choice studies utilized the control function approach introduced by Petrin & Train (2010). For example, Malone & Lusk (2018b) used the control function approach to distinguish perceptions from preferences for beer brands. The authors identified perceptions about other brands as instruments to explain perceptions of the studied brand. They found that perceived taste and brand familiarity are key determinants of choice. Other food choice studies have used hybrid choice models (HCMs), which were introduced by Ben-Akiva et al. (2002) to account for potential endogeneity problems.[15] For instance, Yangui et al. (2016) employed the HCM to investigate the effect of personality traits, lifestyles, and purchasing habits on consumer purchase behavior regarding extra virgin olive oil. Results from this study highlight the importance of behavioral factors for policy design and implementation.

[15]The HCM contains observed choices and additional sets of exploratory variables or latent constructs representing behavioral factors. This allows researchers to simultaneously estimate coefficients for the structural equation, the measurement equation, and the choice model components (see Ben-Akiva et al. 2002, Bolduc et al. 2005).

An HCM approach has also been proposed by Hess & Hensher (2013) to model self-reported ANA statements into discrete choice models. The authors addressed endogeneity in ANA by introducing a hybrid model that simultaneously fits socioeconomic covariates to ANA variables. We argue that the socioeconomic covariates used in this manner would not be a defensible instrument because they would not satisfy the exclusion restriction. For instance, why should a person with a given socioeconomic profile systematically engage in a different form of ANA from others with different socioeconomic profiles? Instead, their inclusion is expected to reduce heterogeneity, as socioeconomic covariates are typically used to condition at least part of the unobservable affecting preference variability. As a result, endogeneity correction can be confused with heterogeneity correction at a high computational cost and requiring substantial coding effort. Furthermore, HCMs require the simultaneous satisfaction of several specification assumptions and generate substantially lower sample likelihoods. Further still, previous studies using the HCM approach to model ANA behavior have found only a modest impact on implied WTP estimates. Examples include studies by Hess & Hensher (2013) and Bello & Abdulai (2016). The latter study uses this approach to study Nigerian consumer preferences for organic food. The authors report no substantial differences in WTP estimates between the hybrid and standard ANA model when the percentages of the self-reported ANA are quite high across attributes. Alternative solutions with the potential to address endogeneity issues in choice modeling are still unexplored in the food choice literature. These include the use of attention allocation functions (Cameron & De Shazo 2010) or context function for shrinkage factors (separately introduced in different forms by Collins 2012 and Hole 2011). All of these, however, are options involving some caveats and, so far, none is emerging as generally accepted as superior. There is, therefore, a definite need for further research in this area.

6. FOOD POLICY EVALUATIONS AND DISCRETE CHOICE EXPERIMENTS

The DCE method is also emerging as a valid tool to evaluate ex ante the effectiveness of food policies. Early research in this domain has focused on evaluating various nutritional food labeling programs, including front-of-package nutritional labels and health claims (Van Wezemael et al. 2014, Thiene et al. 2018). For example, Thiene et al. (2018) looked at different formats of nutritional signals in front-of-package labels of food baskets, linking preference heterogeneity to age groups and self-reported body weight measures, whereas Van Wezemael et al. (2014) conducted a multicountry DCE study on lean beef steak selection to investigate whether front-of-package nutritional labels and health claims are appealing to European consumers.

More recent applications evaluated the effects of fiscal policies on choice outcomes and welfare estimates. For example, Papoutsi et al. (2015) conducted a laboratory-based R-DCE to examine the effects of both a fat tax to discourage unhealthy diets and a subsidy to encourage a healthy diet. The study included a control and both within- and between-subject experiments depending on the price variations induced by the fiscal policies (fat tax and/or subsidy), the decision environment (pestering or pestering), and information provision (information about the fiscal policy intervention or no information). The authors found that the strongest shifts in parental choices toward healthier food products occur when the fat tax and food subsidy are implemented simultaneously, although child pestering can reduce the effect the intervention has on parent's food choices. Ahn & Lusk (2021) used an online DCE to examine the size and direction of the nonpecuniary effect of sugar-sweetened beverage taxes and a size ban. In the experiment, 1,300 US consumers were asked to make beverage choices before and after a price increase or before and after a beverage volume (size) decrease (within-subject comparison) with alternative explanations for the change (between-subject comparison). The design employed by the authors enabled them to compare the difference in sugar-sweetened soda market share before and after a policy across information treatments by exogenously changing the prices or size. Data collected in 2016 showed significant nonpecuniary effects on soda choices, but a replication study in 2019 was unable to find consistent nonpecuniary effects, and the data generally displayed substantial heterogeneity across respondents in the difference-in-difference estimates.

However, the use of DCEs in policy evaluations remains limited. The most limiting factor in standard food DCEs is that respondents are typically asked to only select one of the experimentally designed product alternatives (the preferred one). As illustrated by Caputo & Lusk (2022), this implies that product alternatives are forced to be demand substitutes, and cross-price elasticities are forced to be positive (see Louviere et al. 2000, Hensher et al. 2015). Therefore, if we implement a food policy that is targeted toward meat, for example, we might also affect the demand for pasta or salad. Those relationships, however, fail to be captured by standard DCEs. Caputo & Lusk (2022) propose a basket-based choice experiment approach, which allows researchers to capture both substitution and complementarity patterns owing to its increased flexibility. The authors show that the basket-based choice experiment is an ideal method to measure various food-related policies, with the potential to contribute to the discussion of broader societal significance.

Another limiting factor in food DCEs is that the product quantity is often predefined by the researcher and held constant across choice questions. For example, in typical food DCEs respondents are asked to select their preferred product alternative, but they cannot freely choose the desired consumption level for the selected alternative (Corsi 2007). This obviously prevents researchers from capturing individual satiating behavior, which may have a pivotal role in the study of nutrition and health-related food policies. Corsi (2007) presents alternative theoretical and econometric approaches to incorporate quantity in nonmarket valuation studies. Open-ended and closed-ended formats are discussed by the author as possible methods to incorporate choice quantity in nonmarket valuation studies. A recent food DCE study by Dennis et al. (2021) introduced an open-ended DCE method, called an OECE, and compared it to a standard DCE. Using meat selection as the empirical application, the authors asked survey respondents to select their preferred meat alternative (same weight) or none of them in the standard DCE. In the OECE survey, respondents were presented with different quantity levels and asked to select the number of pounds of each meat product they would purchase, if any, with the selection of zero pounds considered as the no-purchase option. Findings from this study highlight the importance of incorporating quantity as an explicit aspect of the decision-making process in food DCEs. Recent modeling developments allow researchers to model food basket choices by also accounting for satiation effects. Examples include the multiple discrete continuous extreme value model proposed by Bhat (2005, 2008).

Further, most food DCEs are conducted at only one point in time, preventing the study of the evolution (i.e., the dynamics) of preferences. The FooDS series developed by Lusk (2017) is the only research program that uses food DCEs to track consumer preferences over time, enabling researchers to test preference stability over time and possibly its dynamics. Indeed, with the FooDS study focusing on meat selection, the results clearly show that consumer preferences change over time owing to various internal and external factors (Lusk & Tonsor 2016, Lusk 2017). Still, despite its broader societal impact, FooDS remains one of the few studies exploring consumer dynamics using DCEs.

7. FINAL REMARKS AND CONCLUSION

We reviewed the recent methodological progress in food DCEs. We note that despite the substantive progress made, especially in survey design and data modeling, many important conceptual and methodological questions remain unresolved. We hence conclude by making some recommendations and suggesting some promising directions. These could be embraced by future empirical and methodological research to provide much-needed, policy-relevant evidence.

Our first recommendation pertains to ex ante statistical power analysis for treatment effects. Although currently underdeveloped, at least in the context of food DCEs, power analysis can guide the selection of the experimental design and the sampling strategy (Palm-Forster & Messer 2021). This could also be coupled with the use of hybrid experimental designs, where designs are derived to address more than one issue, for example, WTP estimation and prediction. Second, most food choice studies are based on random utility theory. Other theoretical choice frameworks should be explored, such as those based on regret minimization (Chorus 2010) that may be particularly useful when looking at food choices of those concerned with health and nutrition outcomes. Regret minimization may be applied either on its own or simultaneously with random utility models by using finite mixtures, as done by Boeri et al. (2014). Another promising but understudied method for future food choice research concerns the conceptual models based on resolvable uncertainty introduced by Manski (2004) and implemented by Blass et al. (2010). This approach can be implemented by eliciting consumer's probability of selecting each alternative in the choice set, rather than the preferred choice. This approach was implemented in olive oil choice analysis by Scarpa et al. (2021a) and can be useful to separate the effects of the information sets that characterize hypothetical and real choices. The data collected in this context are amenable to robust estimation practices and may be attractive even in small sample contexts, although they require binary decisions.

There is also a broad scope for more DCEs directed to specific food policy evaluations. The food basket approach proposed by Caputo & Lusk (2022) could be extended to also include other choice determinants (quantity) and food choice environments (food consumed away from home). This would allow researchers to create data sets that mimic the type of revealed preference data (e.g., scanner data) more closely. With data from more realistic choice contexts, researchers will be able to evaluate policies by looking specifically at relationships of substitution/complementarity within food choice patterns and at satiation behavior, while avoiding endogeneity issues. As mentioned earlier in this review, this new line of research could also be supported by recent developments in the random utility modeling of multiple discrete-continuous data by means of extreme value analysis, as developed by Bhat (2005, 2008). Future food policy–based research should also focus on determining the effects of internal and external reference prices on fiscal policy evaluations (e.g., soda taxes). For example, prior research shows poor consumer knowledge on soda taxes, which may be driven by high levels of uncertainty about market price distributions (for a discussion, see Caputo & Just 2022).

Finally, while the studies and findings discussed in this review are mostly focused on food DCEs, we are hopeful that the careful reader from other fields interested in DCE applications will find inspiration and practical insights that are transferable to their respective fields of study.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was partly funded by the US Department of Agriculture (USDA) National Institute of Food and Agriculture, Hatch Project 1013332. We are grateful to Jinhua Zhao for detailed feedback on a previous draft. We would also like to thank the Annual Reviews staff for their valuable inputs.

LITERATURE CITED

Ahn S, Lusk JL. 2021. Non-pecuniary effects of sugar-sweetened beverage policies. *Am. J. Agric. Econ.* 103(1):53–69

Alemu MH, Olsen SB. 2018. Can a Repeated Opt-Out Reminder remove hypothetical bias in discrete choice experiments? *Eur. Rev. Agric. Econ.* 45(5):749–82

Alfnes F, Guttormsen AG, Steine G, Kolstad K. 2006. Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives. *Am. J. Agric. Econ.* 88(4):1050–61

Aprile MC, Caputo V, Nayga RM. 2012. Consumers' valuation of food quality labels: the case of the European geographic indication and organic farming labels. *Int. J. Consum. Stud.* 36(2):158–65

Aravena C, Martinsson P, Scarpa R. 2014. Does money talk? The effect of a monetary attribute on the marginal values in a choice experiment. *Energy Econ.* 44:483–91

Arentze T, Borgers A, Timmermans H, DelMistro R. 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transp. Res. E Logist. Transp. Rev.* 39(3):229–44

Atzori R, Pellegrini A, Lombardi GV, Scarpa R. 2022. Response-times and subjective complexity of food choices: a web-based experiment across 3 countries. *Soc. Sci. Comput. Rev.* In press

Balcombe K, Bitzios M, Fraser I, Haddock-Fraser J. 2014. Using attribute importance rankings within discrete choice experiments: an application to valuing bread attributes. *J. Agric. Econ.* 65(2):446–62

Balcombe K, Chalak A, Fraser I. 2009. Model selection for the mixed logit with Bayesian estimation. *J. Environ. Econ. Manag.* 57(2):226–37

Balcombe K, Fraser I, Mcsorley E. 2015. Visual attention and attribute attendance in multi-attribute choice experiments. *J. Appl. Econom.* 30(3):447–67

Bansal P, Daziano RA, Achtnicht M. 2018. Extending the logit-mixed logit model for a combination of random and fixed parameters. *J. Choice Model.* 27:88–96

Bateman IJ, Day BH, Jones AP, Jude S. 2009. Reducing gain-loss asymmetry: a virtual reality choice experiment valuing land use change. *J. Environ. Econ. Manag.* 58(1):106–18

Bazzani C, Caputo V, Nayga RM, Canavari M. 2017a. Revisiting consumers' valuation for local versus organic food using a non-hypothetical choice experiment: Does personality matter? *Food Qual. Prefer.* 62:144–54

Bazzani C, Caputo V, Nayga RM, Canavari M. 2017b. Testing commitment cost theory in choice experiments. *Econ. Inq.* 55(1):383–96

Bazzani C, Palma MA, Nayga RM. 2018. On the use of flexible mixing distributions in WTP space: an induced value choice experiment. *Aust. J. Agric. Resour. Econ.* 62(2):185–98

Bech M, Kjaer T, Lauridsen J. 2011. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Econ.* 20(3):273–86

Bello M, Abdulai A. 2016. Impact of ex-ante hypothetical bias mitigation methods on attribute non-attendance in choice experiments. *Am. J. Agric. Econ.* 98(5):1486–506

Ben-Akiva M, McFadden D, Train K, Walker J, Bhat C, et al. 2002. Hybrid choice models: progress and challenges. *Mark. Lett.* 13(3):163–75

Bhat CR. 2005. A multiple discrete-continuous extreme value model: formulation and application to discretionary time-use decisions. *Transp. Res. B Methodol.* 39(8):679–707

Bhat CR. 2008. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transp. Res. B Methodol.* 42(3):274–303

Blass AA, Lach S, Manski CF. 2010. Using elicited choice probabilities to estimate random utility models: preferences for electricity reliability. *Int. Econ. Rev.* 51(2):421–40

Bliemer MCJ, Rose JM. 2005. *Efficiency and sample size requirements for stated choice studies*. Work. Pap., ITLS, Univ. Sydney, Aust.

Boeri M, Scarpa R, Chorus CG. 2014. Stated choices and benefit estimates in the context of traffic calming schemes: utility maximization, regret minimization, or both? *Transp. Res. A Policy Pract.* 61:121–35

Bolduc D, Ben-Akiva M, Walker J, Michaud A. 2005. Hybrid choice models with logit kernel: applicability to large scale models. In *Integrated Land-Use and Transportation Models*, ed. MEH Lee-Gosselin, ST Doherty, pp. 275–302. Bingley, UK: Emerald Group Publ.

Brooks K, Lusk JL. 2010. Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *Am. J. Agric. Econ.* 92(4):1229–41

Cameron TA, DeShazo JR. 2010. Differential attention to attributes in utility-theoretic choice models. *J. Choice Model.* 3(3):73–115

Cameron TA, Poe GL, Ethier RG, Schulze WD. 2002. Alternative non-market value-elicitation methods: Are the underlying preferences the same? *J. Environ. Econ. Manag.* 44(3):391–425

Campbell D, Hutchinson WG, Scarpa R. 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environ. Resour. Econ.* 41(3):401–17

Campbell D, Hutchinson WG, Scarpa R. 2009. Using choice experiments to explore the spatial distribution of willingness to pay for rural landscape improvements. *Environ. Plan. A* 41(1):97–111

Campbell D, Mørkbak MR, Olsen SB. 2018. The link between response time and preference, variance and processing heterogeneity in stated choice experiments. *J. Environ. Econ. Manag.* 88:18–34

Caputo V. 2020. Does information on food safety affect consumers' acceptance of new food technologies? The case of irradiated beef in South Korea under a new labelling system and across different information regimes. *Aust. J. Agric. Resour. Econ.* 64(4):1003–33

Caputo V, Just D. 2022. The economics of food related policies: considering public health and malnutrition. In *Handbook of Agricultural Economics*, Vol. 6, ed. CB Barrett, DR Just. New York: Elsevier

Caputo V, Lusk JL. 2020. What agricultural and food policies do U.S. consumer prefer? A best-worst scaling approach. *Agric. Econ.* 51(1):75–93

Caputo V, Lusk JL. 2022. A basket-based choice experiment: a method for food demand policy analysis. *Food Policy* 109:102252

Caputo V, Lusk JL, Nayga RM. 2018a. Choice experiments are not conducted in a vacuum: the effects of external price information on choice behavior. *J. Econ. Behav. Organ.* 145:335–51

Caputo V, Lusk JL, Nayga RM. 2020. Am I getting a good deal? Reference-dependent decision making when the reference price is uncertain. *Am. J. Agric. Econ.* 102(1):132–53

Caputo V, Nayga RM, Scarpa R. 2013a. Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Aust. J. Agric. Resour. Econ.* 57(4):465–82

Caputo V, Scarpa R, Nayga RM. 2017. Cue versus independent food attributes: the effect of adding attributes in choice experiments. *Eur. Rev. Agric. Econ.* 44(2):465–82

Caputo V, Scarpa R, Nayga RM, Ortega DL. 2018b. Are preferences for food quality attributes really normally distributed? An analysis using flexible mixing distributions. *J. Choice Model.* 28:10–27

Caputo V, Van Loo EJ, Scarpa R, Nayga RM, Verbeke W. 2018c. Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments. *J. Agric. Econ.* 69(1):35–57

Caputo V, Vassilopoulos A, Nayga RM, Canavari M. 2013b. Welfare effects of food miles labels. *J. Consum. Aff.* 47(2):311–27

Carlsson F, Frykblom P, Lagerkvist JC. 2005. Using cheap talk as a test of validity in choice experiments. *Econ. Lett.* 89(2):147–52

Carlsson F, Martinsson P. 2001. Do hypothetical and actual marginal willingness to pay differ in choice experiments? *J. Environ. Econ. Manag.* 41(2):179–92

Carlsson F, Martinsson P. 2008. How much is too much? An investigation of the effect of the number of choice sets, context dependence and the choice of bid vectors in choice experiments. *Environ. Resour. Econ.* 40(2):165–76

Carson KS, Chilton SM, Hutchinson WG, Scarpa R. 2020. Public resource allocation, strategic behavior, and status quo bias in choice experiments. *Public Choice* 185(1–2):1–19

Carson RT, Groves T. 2007. Incentive and informational properties of preference questions. *Envir. Resour. Econ.* 37:181–210

Caussade S, de Dios Ortúzar J, Rizzi LI, Hensher DA. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transp. Res. B Methodol.* 39(7):621–40

Cerroni S, Watson V, Kalentakis D, Macdiarmid JI. 2019. Value-elicitation and value-formation properties of discrete choice experiment and experimental auctions. *Eur. Rev. Agric. Econ.* 46(1):3–27

Chang JB, Lusk JL, Norwood BF. 2009. How closely do hypothetical surveys and laboratory experiments predict field behavior? *Am. J. Agric. Econ.* 91(2):518–34

ChoiceMetrics. 2021. *Ngene 1.3 User Manual and Reference Guide. The Cutting Edge in Experimental Design*. Sydney, Aust.: ChoiceMetrics. **http://www.choice-metrics.com/NgeneManual130.pdf**

Chorus CG. 2010. A new model of random regret minimization. *Eur. J. Transp. Infrastruct. Res.* 10(2):181–96

Collins AT. 2012. *Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity*. PhD Diss., Inst. Transp. Logist. Stud., Bus. Sch., Univ. Sydney, Aust.

Collins JP, Vossler CA. 2009. Incentive compatibility tests of choice experiment value elicitation questions. *J. Environ. Econ. Manag.* 58(2):226–35

Corsi A. 2007. Ambiguity of measured WTP for quality improvements when quantity is unconstrained: a note. *Eur. Rev. Agric. Econ.* 34(4):501–15

Crastes dit Sourd R. 2021. *A new shifted log-normal distribution for mitigating 'exploding' implicit prices in mixed multinomial logit models*. Work. Pap., Div. Manag., Leeds Univ., UK

Cummings RG, Taylor LO. 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am. Econ. Rev.* 89(3):649–65

Daly A, Hess S, Train K. 2012. Assuring finite moments for willingness to pay in random coefficient models. *Transportation* 39(1):19–31

Darby K, Batte MT, Ernst S, Roe B. 2008. Decomposing local: a conjoint analysis of locally produced foods. *Am. J. Agric. Econ.* 90(2):476–86

Day B, Bateman IJ, Carson RT, Dupont D, Louviere JJ, et al. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. *J. Environ. Econ. Manag.* 63(1):73–91

De-Magistris T, Gracia A, Nayga RM. 2013. On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *Am. J. Agric. Econ.* 95(5):1136–154

De-Magistris T, Pascucci S. 2014. The effect of the solemn oath script in hypothetical choice experiment survey: a pilot study. *Econ. Lett.* 123(2):252–55

De Marchi E, Caputo V, Nayga RM, Banterle A. 2016. Time preferences and food choices: evidence from a choice experiment. *Food Policy* 62:99–109

Dennis EJ, Tonsor GT, Lusk JL. 2021. Choosing quantities impacts individuals' choice, rationality, and willingness to pay estimates. *Agric. Econ.* 52(6):945–62

DeShazo JR, Fermo G. 2002. Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *J. Environ. Econ. Manag.* 44(1):123–43

Fang D, Nayga RM, West GH, Bazzani C, Yang W, et al. 2021. On the use of virtual reality in mitigating hypothetical bias in choice experiments. *Am. J. Agric. Econ.* 103(1):142–61

Fox JA, Shogren JF, Hayes DJ, Kliebenstein JB. 1998. CVM-X: calibrating contingent values with experimental auction markets. *Am. J. Agric. Econ.* 80(3):455–65

Gao Z, Schroeder TC. 2009. Effects of label information on consumer willingness-to-pay for food attributes. *Am. J. Agric. Econ.* 91(3):759–809

Giergiczny M, Valasiuk S, Czajkowski M, De Salvo M, Signorello G. 2012. Including cost income ratio into utility function as a way of dealing with 'exploding' implicit prices in mixed logit models. *J. For. Econ.* 18(4):370–80

Gilbride TJ, Allenby GM. 2004. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Mark. Sci.* 23(3):275–467

Glenk K, Meyerhoff J, Akaichi F, Martin-Ortega J. 2019. Revisiting cost vector effects in discrete choice experiments. *Resour. Energy Econ.* 57:135–55

Grebitus C, Lusk JL, Nayga RM. 2013a. Explaining differences in real and hypothetical experimental auctions and choice experiments with personality. *J. Econ. Psychol.* 36:11–26

Grebitus C, Steiner B, Veeman M. 2013b. Personal values and decision making: evidence from environmental footprint labeling in Canada. *Am. J. Agric. Econ.* 95(2):397–403

Hanley N, Mourato S, Wright RE. 2001. Choice modelling approaches: a superior alternative for environmental valuation? *J. Econ. Surv.* 15(3):435–62

Hensher D, Louviere J, Swait J. 1998. Combining sources of preference data. *J. Econom.* 89(1–2):197–221

Hensher DA. 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environ. Resour. Econ.* 34(1):7–44

Hensher DA, Greene WH. 2003. The mixed logit model: the state of practice. *Transportation* 30(2):133–76

Hensher DA, Rose JM, Greene WH. 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation* 39(2):235–45

Hensher DA, Rose JM, Greene WH. 2015. *Applied Choice Analysis*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.

Hensher DA, Stopher PR, Louviere JJ. 2001. An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application. *J. Air Transp. Manag.* 7(6):373–79

Hess S, Hensher DA. 2013. Making use of respondent reported processing information to understand attribute importance: a latent variable scaling approach. *Transportation* 40(2):397–412

Hole AR. 2011. A discrete choice model with endogenous attribute attendance. *Econ. Lett.* 110(3):203–5

Jacquemet N, Joule RV, Luchini S, Shogren JF. 2013. Preference elicitation under oath. *J. Environ. Econ. Manag.* 65(1):110–32

Johnston RJ, Boyle KJ, Adamowicz W, Bennett J, Brouwer R, et al. 2017. Contemporary guidance for stated preference studies. *J. Assoc. Environ. Resour. Econ.* 4(2):319–405

Kaye-Blake WH, Abell WL, Zellman E. 2009. Respondents' ignoring of attribute information in a choice modelling survey. *Aust. J. Agric. Resour. Econ.* 53(4):547–64

Kessels R, Goos P, Vandebroek M. 2006. A comparison of criteria to design efficient choice experiments. *J. Consum. Res.* 43(3):409–19

Kessels R, Jones B, Goos P, Vandebroek M. 2009. An efficient algorithm for constructing Bayesian optimal choice designs. *J. Bus. Econ. Stat.* 27(2):279–91

Kilders V, Caputo V. 2021. Is animal welfare promoting hornless cattle? Assessing consumer's valuation for milk from gene-edited cows under different information regimes. *J. Agric. Econ.* 72(3):375–59

Kling CL, Phaneuf DJ, Zhao J. 2012. From Exxon to BP: Has some number become better than no number? *J. Econ. Perspect.* 26(4):3–26

Kragt ME. 2013. The effects of changing cost vectors on choices and scale heterogeneity. *Environ. Resour. Econ.* 54(2):201–21

Lancsar E, Swait J. 2014. Reconceptualizing the external validity of discrete choice experiments. *Pharmacoeconomics* 32(10):951–65

Lewis KE, Grebitus C, Nayga RM. 2016. U.S. consumers' preferences for imported and genetically modified sugar: examining policy consequentiality in a choice experiment. *J. Behav. Exp. Econ.* 65:1–8

Lin W, Ortega DL, Caputo V. 2019. Are ex-ante hypothetical bias calibration methods context dependent? Evidence from online food shoppers in China. *J. Consum. Aff.* 53(2):520–44

List JA, Gallet CA. 2001. What experimental protocol influence disparities between actual and hypothetical stated values? *Environ. Resour. Econ.* 20(3):241–54

Loomis J, Gonzalez-Caban A, Gregory R. 1994. Do reminders of substitutes and budget constraints influence contingent valuation estimates? *Land Econ*. 70(4):499–506

Loureiro ML, Lotade J. 2005. Do fair trade and eco-labels in coffee wake up the consumer conscience? *Ecol. Econ.* 53(1):129–38

Loureiro ML, Umberger WJ. 2007. A choice experiment model for beef: what US consumer responses tell us about relative preferences for food safety, country-of-origin labeling and traceability. *Food Policy* 32(4):496–514

Louviere JJ. 2004. *Random utility theory-based stated preference elicitation methods: applications in health economics with special reference to combining sources of preference data*. Work. Pap., Cent. Study Choice 04-001, Univ. Technol., Sydney, Aust.

Louviere JJ, Hensher DA, Swait J. 2000. *Stated Choice Methods: Analysis and Application*. Cambridge, UK: Cambridge Univ. Press

Luchini S, Watson V. 2014. Are choice experiments reliable? Evidence from the lab. *Econ Lett*. 124(1):9–13

Lusk JL. 2003. Using experimental auctions for marketing applications: a discussion. *J. Agric. Appl. Econ.* 35(2):349–60

Lusk JL. 2017. Consumer research with big data: applications from the food demand survey (FooDS). *Am. J. Agric. Econ.* 99(2):303–20

Lusk JL, Norwood FB. 2005. Effect of experimental design on choice-based conjoint valuation estimates. *Am. J. Agric. Econ.* 87(3):771–85

Lusk JL, Norwood FB. 2009. An inferred valuation method. *Land Econ.* 85(3):500–14
Lusk JL, Schroeder TC. 2004. Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *Am. J. Agric. Econ.* 86(2):467–82
Lusk JL, Schroeder TC, Tonsor GT. 2014. Distinguishing beliefs from preferences in food choice. *Eur. Rev. Agric. Econ.* 41(4):27–655
Lusk JL, Shogren J. 2007. *Experimental Auctions: Methods and Applications in Economic and Marketing Research*. Cambridge, UK: Cambridge Univ. Press
Lusk JL, Tonsor GT. 2016. How meat demand elasticities vary with price, income, and product category. *Appl. Econ. Perspect. Policy* 38(4):673–711
Malone T, Lusk JL. 2018a. A simple diagnostic measure of inattention bias in discrete choice models. *Eur. Rev. Agric. Econ.* 45(3):455–62
Malone T, Lusk JL. 2018b. An instrumental variable approach to distinguishing perceptions from preferences for beer brands. *Manag. Decis. Econ.* 39(4):403–17
Manski CF. 1999. Analysis of choice expectations in incomplete scenarios. *J. Risk Uncertain.* 19:49–66
Manski CF. 2004. Measuring expectations. *Econometrica* 72(5):1329–76
Mariel P, Hoyos D, Meyerhoff J, Czajkowski M, Dekker T, et al. 2021. *Environmental Valuation with Discrete Choice Experiments: Guidance on Design, Implementation and Data Analysis*. Cham, Switz.: Springer Nature
Matthews Y, Scarpa R, Marsh D. 2017. Using virtual environments to improve the realism of choice experiments: a case study about coastal erosion management. *J. Environ. Econ. Manag.* 81:93–208
Menapace L, Colson G, Grebitus C, Facendola M. 2011. Consumers' preferences for geographical origin labels: evidence from the Canadian olive oil market. *Eur. Rev. Agric. Econ.* 38(2):193–212
Meyerhoff J, Oehlmann M, Weller P. 2015. The influence of design dimensions on stated choices in an environmental context. *Environ. Resour. Econ.* 61(3):385–407
Mørkbak MR, Christensen T, Gyrd-Hansen D. 2010. Choke price bias in choice experiments. *Environ. Resour. Econ.* 45(4):437–551
Morey ER, Sharma VR, Karlstrom A. 2003. A simple method of incorporating income effects into logit and nested-logit models: theory and application. *Am. J. Agric. Econ.* 85:248–53
Moser R, Raffaelli R, Notaro S. 2014. Testing hypothetical bias with a real choice experiment using respondents' own money. *Eur. Rev. Agric. Econ.* 41(1):25–46
Murphy JJ, Allen PG, Stevens TH, Weatherhead D. 2005. A meta-analysis of hypothetical bias in stated preference valuation. *Environ. Resour. Econ.* 30(3):313–25
Oehlmann M, Meyerhoff J, Mariel P, Weller P. 2017. Uncovering context-induced status quo effects in choice experiments. *J. Environ. Econ. Manag.* 81:59–73
Onozaka Y, McFadden DT. 2011. Does local labeling complement or compete with other sustainable labels? A conjoint analysis of direct and joint values for fresh produce claim. *Am. J. Agric. Econ.* 93(3):689–702
Palm-Forster L, Messer K. 2021. Experimental and behavioral economics to inform agri-environmental program and policies. In *Handbook of Agricultural Economics*, Vol. 5, ed. CB Barrett, DR Just, pp. 4331–4406. New York: Elsevier
Palma MA. 2022. Psychosociological measures and consumer food choice. In *Handbook of Agricultural Economics*, Vol. 6, ed. CB Barrett, DR Just. New York: Elsevier
Palma MA, Vedenov DV, Bessler D. 2020. The order of variables, simulation noise, and accuracy of mixed logit estimates. *Empiric. Econ.* 58:2049–83
Papoutsi GS, Nayga RM, Lazaridis P, Drichoutis AC. 2015. Fat tax, subsidy or both? The role of information and children's pester power in food choice. *J. Econ. Behav.* 117:196–208
Payne JW, Bettman JR, Johnson EJ. 1992. Behavioral decision research: a constructive processing perspective. *Annu. Rev. Psychol.* 43:87–131
Penn JM, Hu W. 2018. Understanding hypothetical bias: an enhanced meta-analysis. *Am. J. Agric. Econ.* 100(4):1186–1206
Petrin A, Train K. 2010. A control function approach to endogeneity in consumer choice models. *J. Mark. Res.* 47:3–13
Ready RC, Champ PA, Lawton JL. 2010. Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. *Land Econ.* 86:363–68

Revelt D, Train K. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Rev. Econ. Stat.* 80(4):647–57

Roe BE, Just DR. 2009. Internal and external validity in economics research: tradeoffs between experiments, field experiments, natural experiments, and field data. *Am. J. Agric. Econ.* 91(5):1266–71

Rose JM, Hensher DA, Caussade S, de Dios Ortúzar J, Jou RC. 2009. Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. *J. Transp. Geogr.* 17(1):21–29

Rose JM, Masiero L. 2010. A comparison of the impacts of aspects of prospect theory on WTP/WTAEstimated in preference and WTP/WTA space. *Eur. J. Transp. Infrastruct. Res.* 10(4). **https://journals.open.tudelft.nl/ejtir/article/view/2898**

Sándor Z, Wedel M. 2001. Designing conjoint choice experiments using managers' prior beliefs. *J. Mark. Res.* 38(4):430–44

Sarrias M, Daziano RA. 2018. Individual-specific point and interval conditional estimates of latent class logit parameters. *J. Choice Model.* 27:50–61

Scarpa R, Bazzani C, Begalli D, Capitello R. 2021a. Resolvable and near-epistemic uncertainty in stated preference for olive oil: an empirical exploration. *J. Agric. Econ.* 72(2):335–69

Scarpa R, Campbell D, Hutchinson WG. 2007. Benefit estimates for landscape improvements: sequential Bayesian design and respondents' rationality in a choice experiment. *Land Econ*. 83(4):617–34

Scarpa R, Del Giudice T. 2004. Market segmentation via mixed logit: extra-virgin olive oil in urban Italy. *J. Agric. Food Ind. Organ.* 2(1):1–20

Scarpa R, Ferrini S, Willis K. 2005. Performance of error component models for status-quo effects in choice experiments. In *Applications of Simulation Methods in Environmental and Resource Economics*, ed. R Scarpa, A Alberini, pp. 247–73. Dordrecht, Neth.: Springer

Scarpa R, Franceschinis C, Thiene M. 2021b. Logit mixed logit under asymmetry and multimodality of WTP: a Monte Carlo evaluation. *Am. J. Agric. Econ.* 103(2):643–62

Scarpa R, Gilbride TJ, Campbell D, Hensher DA. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *Eur. Rev. Agric. Econ.* 36(2):151–74

Scarpa R, Notaro S, Louviere J, Raffaelli R. 2011. Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in alpine grazing commons. *Am. J. Agric. Econ.* 93(3):813–28

Scarpa R, Rose JM. 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Aust. J. Agric. Econ.* 52(3):253–82

Scarpa R, Thiene M, Hensher DA. 2010. Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: Does it matter? *Land Econ*. 86(4):817–39

Scarpa R, Thiene M, Marangon F. 2008. Using flexible taste distributions to value collective reputation for environmentally friendly production methods. *Can. J. Agric. Econ.* 56(2):145–62

Scarpa R, Zanoli R, Bruschi V, Naspetti S. 2013. Inferred and stated attribute non-attendance in food choice experiments. *Am. J. Agric. Econ.* 95(1):165–80

Simon HA. 1955. A behavioral model of rational choice. *Q. J. Econ.* 69(1):98–118

Street DJ, Burgess L, Louviere JJ. 2005. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. *Int. J. Res. Mark.* 22(4):459–70

Swait J, Adamowicz W. 2001. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organ. Behav. Hum. Decis. Process.* 86(2):141–67

Tait P, Saunders C, Dalziel P, Rutherford P, Driver T, Guenther M. 2019. Estimating wine consumer preferences for sustainability attributes: a discrete choice experiment of Californian sauvignon blanc purchasers. *J. Clean. Prod.* 233:412–20

Thiene M, Scarpa R. 2009. Deriving and testing efficient estimates of WTP distributions in destination choice models. *Environ. Resour. Econ.* 44(3):379–91

Thiene M, Scarpa R, Longo A, Hutchinson WG. 2018. Types of front of pack food labels: Do obese consumers care? Evidence from Northern Ireland. *Food Policy* 80:84–102

Train KE. 1998. Recreation demand models with taste differences over people. *Land Econ*. 74(2):230–39

Train KE. 2009. *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge Univ. Press

Train KE 2016. Mixed logit with a flexible mixing distribution. *J. Choice Model.* 19:40–53
Train KE, Weeks M. 2005. Discrete choice models in preference space and willingness-to-pay space. In *Applications of Simulation Methods in Environmental and Resource Economics. The Economics of Non-Market Goods and Resources*, Vol. 6, ed. R Scarpa, A Alberini, pp. 1–16. Dordrecht, Neth.: Springer
Van Loo EJ, Caputo V, Lusk JL. 2020. Consumer preferences for farm-raised meat, lab-grown meat, and plant-based meat alternatives: Does information or brand matter? *Food Policy* 95:101931
Van Loo EJ, Grebitus C, Nayga RM, Verbeke W, Roosen J. 2018a. On the measurement of consumer preferences and food choice behavior: the relation between visual attention and choices. *Appl. Econ. Perspect. Policy* 40(4):539–62
Van Loo EJ, Grebitus C, Roosen J. 2019. Explaining attention and choice for origin labeled cheese by means of consumer ethnocentrism. *Food Qual. Prefer.* 78:103716
Van Loo EJ, Nayga RM, Campbell D, Seo HS, Verbeke W. 2018b. Using eye tracking to account for attribute non-attendance in choice experiments. *Eur. Rev. Agric. Econ.* 45(3):333–65
Van Wezemael L, Caputo V, Nayga RM, Chryssochoidis G, Verbeke W. 2014. European consumer preferences for beef with nutrition and health claims: a multi-country investigation using discrete choice experiments. *Food Policy* 44:167–76
Vermeulen B, Goos P, Scarpa R, Vandebroek M. 2011. Bayesian conjoint choice designs for measuring willingness to pay. *Environ. Resour. Econ.* 48(1):129–49
Vossler CA, Kerkvliet J, Polasky S, Gainutdinova O. 2003. Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon. *J. Econ. Behav. Organ.* 51(2):261–77
Walker JL, Wang Y, Thorhauge M, Ben-Akiva M. 2018. D-efficient or deficient? A robustness analysis of stated choice experimental designs. *Theory Decis.* 84(2):215–38
Yangui A, Costa-Font M, Gil JM. 2016. The effect of personality traits on consumers' preferences for extra virgin olive oil. *Food Qual. Prefer.* 51:27–38
Zhao J, Kling CL. 2004. Willingness to pay, compensating variation, and the cost of commitment. *Econ. Inq.* 42(3):503–17

Annual Review of Resource Economics

Volume 14, 2022

Contents

Environmental Economics

Resource Economics

Energy Economics

Errata

An online log of corrections to *Annual Review of Resource Economics* articles may be found at http://www.annualreviews.org/errata/resource