Food Quality and Preference 102 (2022) 104678

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

The state of the art of discrete choice experiments in food research

Sebastien Lizin [1,*], Sandra Rousseau [1], Roselinde Kessels, Michel Meulders, Guido Pepermans, Stijn Speelman, Martina Vandebroek, Goedele Van Den Broeck, Ellen J. Van Loo, Wim Verbeke

Universiteit Hasselt, Martelarenlaan 42, B-3500 Hasselt, Belgium

ARTICLE INFO

Keywords:
DCE
Food choice
Methodological review
Valuation
Best practice

ABSTRACT

Discrete choice experiments (DCEs) have become an often-used research method in food research due to their ability to uncover trade-offs made when choosing among multiple alternatives, especially when dealing with credence attributes. Insights into the main elements of the consumers' decision-making process are key to informing both public and private policies related to food production and consumption. However, DCEs are not confined to this field of study. This narrative methodological review sets out to provide a critical appraisal of the state of the art of DCEs in food research. We logically structure our review by comparing the field-independent state-of-the-art to its application in the specific food choice research domain. The comparison is presented for each of the steps required in implementing DCEs and allows for the identification of areas of improvement in best practice. We find that food research has adopted many of the methodological advances over the years, but further improvements are encouraged and outlined. Recommendations for future research are discussed.

1. Introduction

1.1. Why food DCEs

Billions of people make dozens of food purchasing and consumption decisions every day. While one part of those decisions might be merely habitual, based on routines informed by favorable past experience and satisfaction or simply guided by low involvement, other food choices may require at least some kind of active reasoning or deliberation (e.g., Bublitz et al., 2010; Gorton & Barjolle, 2013; Nardi et al., 2019). People make food choices in a multitude of choice contexts, combining different moments, occasions, situations, and types of company, and they do so while having heterogeneous sets of personal characteristics, knowledge, beliefs, perceptions, attitudes, and motivations (e.g., Steptoe et al., 1995; Gorton & Barjolle, 2013; Nardi et al., 2019). In contexts where choice is available, alternatives are often plentiful and each alternative food option combines multiple tangible and intangible characteristics or attributes. Since food attributes can have positive or negative utility impacts and can be seen as being more or less important, trade-offs are needed. Moreover, food choices not only have an impact on a person's nutritional and health status and on his/her overall well-being, they also have an impact on our living environment, on social interactions and on society as a whole (e.g., Reisch et al., 2013). As food choices entail personal and societal risks and raise ethical issues, food consumption and production as well as individuals' responsibility therein have become increasingly debated during the past decades (e.g. Dieterle, 2022). This triggers an interest in better understanding people's food choices in a particular context as the key to informing public and private policies (e.g. Reisch et al., 2013; Van Loo et al., 2020). These initiatives range from institutional and governmental policies to private managerial policies and marketing strategies of actors involved in food supply chains, and at scales that extend from local to global. Therefore, it is unsurprising that food choice has emerged as an important application field in the research domain of discrete choice experiments (DCEs).

A DCE is a method of identifying the attributes that drive the preferences of food producers and consumers with respect to a variety of issues described above. Several steps have been identified in the implementation of a DCE (e.g., Ryan et al., 2008; Holmes et al., 2017).

* Corresponding author.
E-mail address: sebastien.lizin@uhasselt.be (S. Lizin).

[1] Sebastien Lizin and Sandra Rousseau share first authorship.

https://doi.org/10.1016/j.foodqual.2022.104678
Received 13 July 2021; Received in revised form 7 April 2022; Accepted 3 July 2022
Available online 7 July 2022

These include problem definition, identification of the attributes and attribute levels, development of the experimental design, survey development, survey implementation, and model estimation, until the interpretation of the results. The latter are utility coefficients that may consequently be converted into other metrics such as choice probabilities, elasticities, or (marginal) willingness-to-pay (WTP) estimates. To obtain these outputs, respondents are given a choice context (such as buying groceries in a supermarket) and asked to choose their preferred alternative or profile out of at least two alternatives, which may be labeled or unlabeled, in a series of choice sets in which the attributes' levels are deliberately varied according to an experimental design. Unlabeled DCEs are typically used to quantify utility coefficients and WTP estimates, whereas labeled DCEs[2] may also be used to derive market shares and elasticities (Louviere et al., 2000). DCEs assume that individuals derive utility from the attributes of the available food options and that individuals' preferences are revealed through their choices (Thurstone, 1931; Lancaster, 1966). DCEs make it possible to infer the value of an attribute from stated or revealed choices, even though the individual may not be aware of this value. This makes a DCE a valuable tool to assess the factors that influence food choices, which are often the results of habits, heuristics, and low involvement decisions.

Next to DCEs, there are other value elicitation methods which can be used in an experimental setting to study consumers' preferences and WTP for food products. Examples are multiple price lists[3] (Asioli et al. 2021), experimental auctions (Canavari et al., 2019), or open-ended choice experiments (Corrigan et al., 2009). We focus on DCEs as it provides a choice setting mimicking the choice situation that consumers generally face in real life (e.g., Louviere et al., 2000). In DCEs, participants are asked to consider several products and select the preferred one. Similarly, when shopping for food in grocery stores or choosing dishes in a restaurant setting, consumers are confronted with a set of possible food options, which vary in attribute levels, and select their preferred option. DCEs therefore allow us to understand current behavior and predict future choices. These features contribute to explaining why the currently available valuation literature in the Web of Science on food valuation is still dominated by DCEs. We refer the reader interested in a more detailed comparison across food valuation methods to recent publications such as Alphonce and Alfnes (2017), Shi et al. (2018), and Asioli et al. (2021). In sum, comparisons reveal that, even in real-contexts, the WTP estimates resulting from a comparative, choice-based elicitation mechanism such as a DCE tend to differ from WTP estimates elicited using non-comparative bids such as resulting from auctions.

1.2. The evolution of the food choice literature

Louviere (1984) was the first to use choice experiments and logit choice models to predict the proportion of consumers willing to try new food products in a fast food restaurant. After that, it took almost two decades before the use of DCEs really took off in the food choice literature. The number of papers reporting on DCEs in food evolved from a single paper in each of 2001 and 2002 to fewer than 10 papers per year until 2008. Since 2009, this number increased slowly to 30 by 2014 and then at a greater pace to reach almost 100 by 2019 and 2020 (see Suppl. Mat.: Fig. 1). This evolution reflects the growing interest in using and reporting DCEs in the food choice domain over the past 20 years, as supported by a growing diversity of journals publishing papers with that type of methodology indexed in the Web of Science (WoS).[4]

Overall, the largest number of DCE papers on food choice addressed topics dealing with food safety or safety risks (n = 188), followed by origin or traceability (n = 172), health or nutrition (n = 129), biotechnology or genetic modification (n = 68) and animal welfare (n = 62). In terms of product categories, the main interest has been in consumer preferences for meat (beef, pork, poultry, processed meat products and, more recently, also for alternatives to conventional meat) (n = 202), followed by organic foods (n = 161), and functional foods or foods with nutrition or health claims (n = 57). Some of the less covered food categories were wine, olive oil, eggs and vegetables. It should be noted that the reported topics and product categorizations are not mutually exclusive, since many studies cover more than one topic and/or product category.

With respect to focal themes, the number of DCE studies on organic production or organic foods has been growing steadily (Fig. 1). Whereas organic production was previously typically considered as the healthier and safer alternative for conventional production, the contextualization changed to organic as the provider of environmental rather than merely health and safety benefits; that is, studying the potential of organic as a more sustainable choice. The evolution of DCE studies on meat shows a more irregular evolution. In 2018, for example, a substantial number of publications focused on safety issues in a Chinese context on one hand, and/or on meat and its eventually more sustainable alternatives on the other hand (e.g. Lai et al., 2018; Wang et al., 2018a). With respect to meat, a gradual shift over time was observed from a focus on meat safety and country-of-origin in the early periods (e.g. Enneking, 2004; Loureiro & Umberger, 2007) to contrasting conventional meat, with more sustainable meat alternatives, more recently (e.g. Slade, 2018; Van Loo et al., 2020). Despite a similar total number of Web of Science published papers using DCEs, fewer papers dealt specifically with biotechnology or animal welfare in 2020 compared to 2019, suggesting a decreasing topicality of these themes most recently. In turn, specific sustainability-related themes, such as food waste reduction and food packaging characteristics, emerged (e.g. Gracia and Gómez, 2020; Wensing et al., 2020).

1.3. The trigger for food choice DCEs

Food production methods are credence attributes that cannot be objectively verified or experienced by consumers (Darby & Karni, 1973). Consumers have to rely on information cues provided, which they may value in case they believe the information and its source are truthful and trustworthy. Efforts to provide food products with such credence attributes are often met with uncertainty and even resistance as the benefits of transforming production processes are uncertain. On one hand, such a

[2] Within the food DCE literature, we only found limited examples of applications of labeled DCEs. Some noteworthy exceptions are the work of Enneking et al. (2007), who combined food DCEs with sensory testing; Nguyen et al. (2015), who investigated preferences for labeled seafoods; and Van Loo et al. (2020), who analyzed consumer preferences for meat and meat alternatives. Ballco and Gracia (2020) provided an overview of previous research that combined intrinsic and extrinsic attributes using conjoint analysis, experimental auctions and DCEs. Note that we have not considered brand choice models, which are typically estimated on scanner panel data, to be DCEs.

[3] Also known as 'payment cards', a specific elicitation format used for contingent valuation.

[4] The main journals publishing DCEs were identified using the Web of Science by performing an unrestricted search on TS=("choice experiment$" AND "food"). This provided 1393 hits on 28 February 2022. This dataset was inspected and the most commonly recurring journals (with 15 or more publications) were then tabulated: *Food Quality and Preference* (5.46 percent), *Sustainability* (3.81 percent), *Food Policy* (3.45 percent), *British Food Journal* (2.94 percent), *Appetite* (2.15 percent), Agribusiness (1.94 percent), *Journal of Agricultural Economics* (1.87 percent), *European Review of Agricultural Economics*, *Journal of Cleaner Production*, *Plos One*, *American Journal of Agricultural Economics*, *Agricultural Economics, Ecological Economics, Canadian Journal of Agricultural Economics - Revue Canadienne d'agroeconomie, Foods,* and *Nutrients*. A similar search in Scopus – using the query (TITLE-ABS-KEY ("choice experiment $" AND food)) - yielded 1311 results with mainly similar journals. Still the following journals with 15 or more publications can be added based on the Scopus search: *Oecologia, Entomologia Experimentalis et Applicata*, *International Food and Agribusiness Management Review* and *Meat Science*.

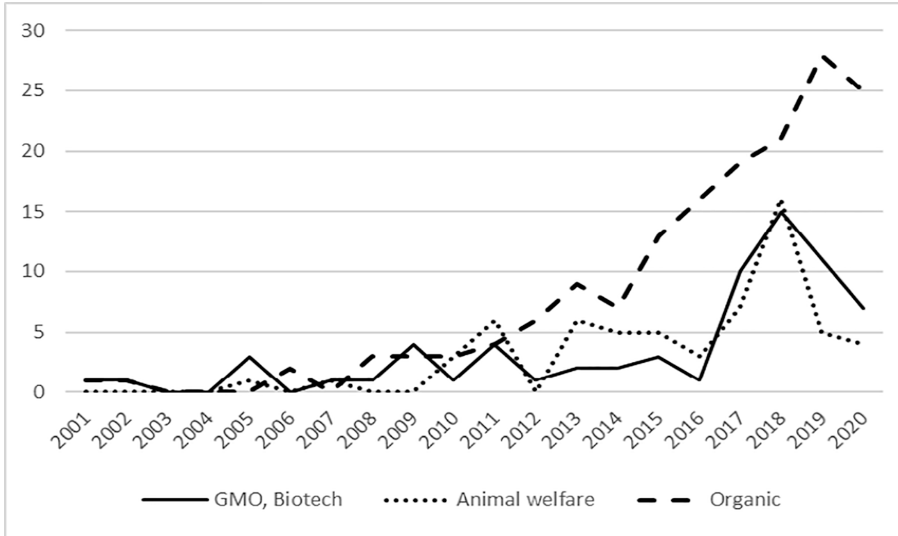

Fig. 1. Number of publications in Web of Science indexed journals using DCEs and focusing on genetically modified organisms or biotechnology, animal welfare or organic as main themes, 2001–2020.

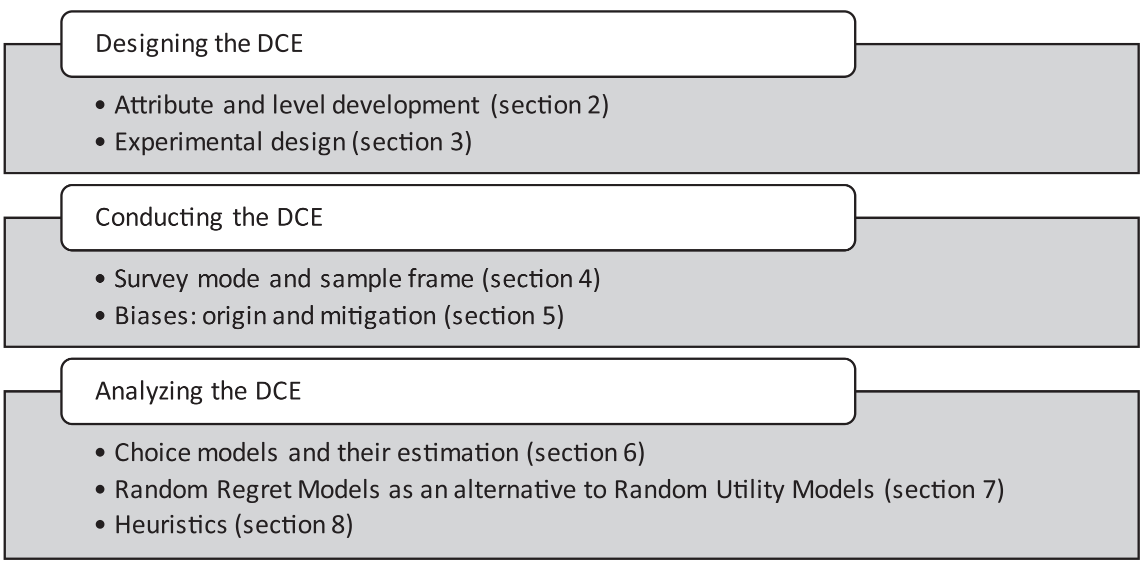

Fig. 2. Structure of the DCE process and paper outline.

transformation often involves the use of new food technologies, unfamiliar ingredients or processing techniques, and is typically more costly than conventional production methods. On the other hand, proving the truth and reliability of production-related claims such as GMO-free production is challenging and can rarely be done with complete certainty. This implies that producers may not reap the benefits from their efforts and that policymakers have difficulty monitoring compliance and measuring the achievement of policy targets. Therefore, it is crucial to understand consumers' reactions to food characteristics and production-related claims or information provisioning and this justifies the interest in assessing willingness-to-accept and willingness-to-pay. Exemplary cases include those of organic production, the presence or absence of GMOs, efforts to improve and signal animal welfare or a product's nutritional value and healthiness, strategies to reduce safety risks and provide related reassurance of guarantees, and novel food production and processing technologies, all of which are communicated to consumers as information alongside the core product or through labels on packaging (e.g., Peschel et al., 2019; Van Loo et al., 2020).

1.4. The contribution and remainder of this paper

The importance of generating valid and reliable insights from DCE studies is an overarching concern. While DCEs can be seen as a flexible and attractive valuation method, their reliability and validity have been questioned; that is, whether they give consistent results across different survey designs that might be used to measure the same quantity (reliability) and whether they measure what they are intended to (validity) (e.g., Bateman et al., 2002; Rakotonarivo et al., 2016; Bishop & Boyle, 2019; Mariel et al., 2021).

In this narrative review we provide an overview of how DCEs are currently used to gain insight into food choices and compare this to general best practice in DCE research. Such initiative can be considered a methodological literature review[5], being "a contribution that formally or informally reviews the existing literature regarding practices about methodological issues, summarizes the literature, and provides recommendations for improved practice" (Aguinis et al., 2020: p2). The latter authors ascribe three potential merits to such papers. First, they may help and guide researchers, including students, doctoral researchers and scholars, to improve their methodological skills. Second, they may contribute to identifying knowledge gaps and research needs. Third, they may be prescriptive in nature and as such describe "how to do things right" as such mitigating questionable research practices. This paper envisages addressing especially the former two as our main goal is to assess where the food DCE subfield leads DCE best-practice and where it is following or lagging. It contributes to science and good practice therein as the exchange of best-practice across fields facilitates the adoption and creation of new knowledge (e.g. Sun & Latora, 2020), which is critical given the sharp rise in food-related research and publications applying DCEs. This contribution provides a synthesis of DCE best-practice across the sequence of steps that compose the DCE methodology and across research fields in view of deriving recommendations related to food choice research.

Consequently, this manuscript is logically structured following the order in which a DCE is respectively designed, conducted, and analyzed. The full sequence of steps is visualized in Fig. 2. After a brief definition of each consecutive step in a DCE, each of the following sections presents the state of the art in general and in food research specifically. We then highlight how the design, implementation, and analysis affects reliability and validity in section 9. We end by formulating methodological recommendations that will extend the best-practice of DCEs for studying food choices.

2. Attribute and level development

It has been argued that "a good DCE is one that has a sufficiently rich set of attributes and choice contexts, together with enough variation in the attribute levels necessary to produce meaningful behavioral responses in the context of the strategies under study" (Ryan et al., 2008, p17). Therefore, the choice of alternatives and their attributes to be considered in the experiment is crucial (e.g., Caussade et al., 2005; Johnston et al., 2017). The alternatives' attributes that are included in the design (see also section 3) explain the observable or systematic part of total utility, whereas unobservable attributes affecting choice are an important cause of unobserved or random variation in preferences. Therefore, the more attributes included in the design, the better the researcher will be able to explain the choices, but the higher the cognitive burden becomes for the respondents (e.g., DeShazo & Fermo, 2002). Hence, the researcher is required to select a limited number of attributes (Green, 1974). However, ignoring important attributes or ambiguously describing them may render them useless for informing policy (e.g., Lancsar & Louviere, 2008; Johnston et al., 2012; Rolfe & Windle, 2015). Hence, the validity of DCEs depends on how complex information about food policies or interventions is transformed into a limited number of relevant attributes.

[5] Our literature review approach consisted of the following steps. The exact keywords used in the Web of Science queries that allowed us to retrieve the majority of the food choice literature mentioned in the remainder of the text is provided in the Suppl. Mat.. After each query, the resulting papers' abstracts were read to verify whether an individual study warranted inclusion. Abstracts were inspected until saturation occurred on a given topic – that is, a specific step in the methodological process of carrying out a DCE – after which the set of literature was synthesized and consequently contrasted with general best practice. As a robustness check we also performed similar searches in Scopus and the results are also reported in the Suppl. Mat..

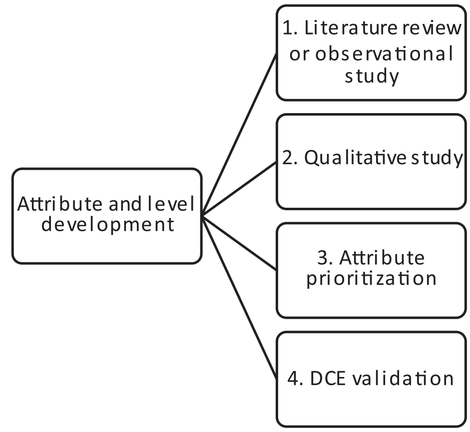

Fig. 3. Attribute and level development as a multi-stage process embedded within setting up DCEs.

Attribute and level development is a multi-stage process (Fig. 3). Given the choice context,[6] a careful selection of core attributes needs to be made before attribute levels are devised that allow the researcher to create an operable DCE. To develop attributes and levels, practitioners have recommended performing a qualitative study based on the results of a (systematic) literature review, an observational study and/or a focus group discussion (e.g., Klojgaard et al., 2012; Helter & Boehler, 2016). This exploratory phase is equally important when the DCE is designed in response to a policy question, as it improves the DCE's content validity (Coast et al., 2012). For a description with regards to how such attribute and level development may be performed, we refer to the Suppl. Material.

Looking at the food related DCE literature, not a single publication that dealt primarily with attribute and level development was identified based on the queries in the Web of Science (see Suppl. Mat.). Hence, the attention devoted to this particular phase in the research set-up has been limited in food research. Moreover, including a description of the process of attribute and level development is given little attention in the more cited literature[7]. Nonetheless, qualitative tools for attribute (level) development have been mentioned more frequently in recent studies. Hence, the process of attribute selection in the food DCE literature has scope for improvement. The identified lack of attention for an elaborate attribute and level development process in food DCEs may result from respondents' general familiarity with the investigated foods, which are often the more popular and widely available products on the market. However, attribute selection becomes more important when the food products are unfamiliar to a larger group of consumers as the econometric estimation also becomes more challenging (e.g., Czajkowski et al., 2015; Heidenreich et al., 2018). However, a search of recent DCE studies on insect-based food products – as an example of an unfamiliar product category – reveals large differences between studies as Videbæk and Grunert (2020) did not provide any information of the attribute selection process, while Alemu et al. (2017) explicitly mentioned an extensive literature review, focus group discussion, and the requirement that the DCE should be credible, realistic, and easy to understand for all participants. Moreover, when applying DCEs in developing countries (see Suppl. Mat.), a number of specific challenges (e.g., Bennett and Birol, 2010) may arise, such as the possible use of alternatives to

[6] Decision mapping can be used to identify distinct choices; see Michaels-Igbokwe et al. (2014).

[7] This does not automatically mean that authors have neglected attribute and level development. It may simply have been omitted due to space limitations and focus on the main research objective(s). Worse than neglect is when interviews and focus groups are mentioned, but give the impression that more was done than what actually occurred.

monetary payment vehicles in low-cash contexts (e.g., Gibson et al., 2016; Vondolia & Navrud, 2019) or the need to keep the choice task simple due to lower education levels (e.g., Gelaw et al., 2016).

3. Experimental design

After selecting attributes and attribute levels, the experimental design, i.e. the selection of profiles into choice sets, is the next focal point of the researcher. To a large extent, the DCE design drives the power of statistical inference and is therefore key in planning a DCE (Hoyos, 2010). Early theoretical design development for DCEs made use of orthogonal level-balanced factorial designs that are commonly associated with linear models. However, most discrete choice models are nonlinear in the parameters, implying that design quality depends on unknown parameters (Sándor & Wedel, 2001). Consequently, researchers need to utilize a priori knowledge about the values for the parameters to generate an efficient design (e.g., Kessels et al., 2008; Bliemer & Collins, 2016; De Marchi et al., 2016).

3.1. Design options

A first possible approach consists of orthogonal factorial designs which assume zero parameter values as prior, meaning that people have no prior preference for any of the attribute levels, which is often not realistic. Still, because these designs are historically rooted in the general design literature (for industrial and agricultural experiments) and well documented in catalogs they are frequently used (e.g., Louviere et al., 2000; Kuhfeld & Tobias, 2005; Street & Burgess, 2007).

Nowadays, thanks to modern technology, a second approach called Bayesian D-optimal design has been developed to fit the choice design problem and is increasingly considered state of the art for DCEs. Bayesian D-optimal designs have most often been generated to precisely estimate the multinomial logit (MNL) model, because they are implemented in statistical software (e.g., Ngene, JMP and the R package idefix) and also perform relatively well in terms of estimating the panel mixed logit (MIXL) model (Bliemer & Rose, 2010). This is convenient because Bayesian D-optimal designs for the MIXL model take longer to generate due to the complexity of the calculations (e.g., Bliemer & Rose, 2010; Traets et al., 2020). More information on generating MNL and MIXL designs can be found in the Suppl. Mat..

Another important category of Bayesian D-optimal designs have been generated based on the no-choice nested logit model. The choice sets in these designs include not only the profiles or real-choice options, but also an opt-out, status-quo or no-choice option (Rousseau, 2015). Such choice sets are particularly valuable if one wants to estimate market shares. Bayesian D-optimal designs involving a no-choice option have been developed for both full profiles (Vermeulen et al., 2008; Goos et al., 2010) and partial profiles (Kessels et al., 2017). They have proven to be more informative for estimating the no-choice nested logit model than the traditional approach of adding a no-choice option to each choice set of a Bayesian D-optimal MNL design that is constructed ignoring the no-choice option. Also worthy of mentioning is the recent introduction of Bayesian D- and I-optimal mixture designs for DCEs, where food products are described as mixtures of ingredients (e.g., Ruseckaite et al., 2017; Goos & Hamidouche, 2019; Becerra & Goos, 2021). These designs are optimized for mixture-choice models where Scheffé mixture models (Scheffé, 1963) replace the systematic utilities of the choice models for these food products.

DCEs on food choices (see Suppl. Mat.) – as with many application fields of DCEs – are gradually adopting a Bayesian D-optimal design approach for the MNL model (e.g., Czine et al., 2020; Paffarini et al., 2021). Its use is often preceded by a pilot survey based on an orthogonal factorial design to obtain the priors for the Bayesian main design (e.g., Scarpa et al., 2013; Zanoli et al., 2013; De Marchi et al., 2016). Such a sequential design strategy is a safe approach, but is not required since one can specify an uninformative prior distribution, like the uniform distribution, for the parameters. Apart from the upsurge of Bayesian D-optimal designs, orthogonal factorial designs are still frequently used (e.g., Caputo et al., 2013; Palma et al., 2018).

3.2. Choice complexity

The validity of a DCE depends not only on its statistical quality as ensured by the experimental design, but also on the choice task complexity (e.g., Johnson et al., 2013). The overall design quality depends on both statistical and response qualities. Out of all the design dimensions – that is, the number of choice sets and profiles in a choice set, the number of attributes and attribute levels, and the range of those levels – the number of attributes has the greatest influence on the error variance (Caussade et al., 2005). Similarly, Meyerhoff et al. (2015) revealed that design dimensions may influence error variance. Respondents can process only a limited number of attributes depending on the application (Green, 1974). To investigate larger numbers of attributes, the levels of only a subset of the attributes in a choice set are varied in so-called partial profile designs (see Suppl. Mat.); these designs contrast with the traditional full-profile designs that allow the levels of all attributes to vary (Green, 1974; Kessels et al., 2011, 2015).

4. Survey mode and sample frame

Obtaining unbiased and consistent DCE results depends heavily on the way the survey is administered and distributed to respondents and how respondents are sampled from the target population (e.g. Brace, 2018). Apart from question formulation, survey length, and the incentives for participation, survey mode and data collection methods are especially worth mentioning. The sample selection procedure has important consequences regarding the representativeness and generalizability of the DCE findings, as well as the cost associated with sampling and data collection. Representativeness is crucial when the researcher wants to provide useful advice to policy makers, organizations, or businesses. A representative dataset can be created by a sufficiently large probabilistic sample (see power calculations, e.g., Dupont & Plummer, 1990, or de Bekker-Grob et al., 2015, specifically for DCE), while a non-probabilistic quota sampling method can generate a dataset that is representative of predetermined, observable characteristics of the target population. However, if these characteristics are not correlated with unobservable preferences, this method will not lead to a representative sample.

There is a range of survey modes to implement a DCE and to assess respondents' preferences (see Suppl. Mat.). Since the 2010 s, internet/e-based technologies have emerged as the most common survey mode, including surveys on online platforms or via e-mail[8] (Lindhjem & Navrud, 2011; Angeliki et al., 2016). Compared to face-to-face interviews and postal/telephone surveys, internet/e-based surveys have the advantage of a relatively low implementation cost and a relatively quick data collection (Olsen, 2009; Windle & Rolfe, 2011). In addition, self-registered web surveys might be less exposed to social desirability bias (see Section 5.2) than interviewer-registered surveys (Kreuter et al., 2008). Within web surveys, the type of device that respondents use (for example, desktop/laptop or mobile devices, such as tablets and smartphones) might influence the results. However, Liebe et al. (2015) found, for a case study of German citizens' preferences for renewable energy, that the use of mobile devices did not affect the tendency to choose the status-quo option nor the scale compared to the use of a desktop/laptop.

An important concern that has arisen with the increased use of internet/e-based surveys is the representativeness of the sample (Boyle

[8] Face-to-face interviews are still the dominant survey mode in low- and middle-income countries (Bennet & Birol, 2010). Especially when collecting data in a remote area setting, face-to-face interviews might be the only option (Liebe et al., 2020).

et al., 2016). Representation error might occur when the sample frame is not representative of the targeted population, due to under-coverage, lower response rates, and more protest bidders (Angeliki et al., 2016). Most of these issues can be mitigated by using well-recruited internet panel samples or sending out personal invitations, if possible coupled with a reward for the respondent (usually some kind of voucher) (Olsen, 2009). However, many studies found similar welfare estimates across different types of survey modes, suggesting that – when correctly implemented - the survey mode does not influence the results of a choice experiment (e.g., Olsen, 2009; Lindhjem & Navrud, 2011; Windle & Rolfe, 2011). For a thorough overview of a correct implementation of stated preference valuation web surveys, we refer to the study by Angeliki et al. (2016).

No papers were retrieved based on the used query that specifically focused on survey mode and sample frame in the context of food-related DCEs. Szolnoki and Hoffmann (2013) compared different survey modes in wine consumption in Germany, but without using a DCE. They found that face-to-face and telephone surveys resulted in the most representative sample, but that web-based surveys (especially in case of snowball sampling) should be corrected using population weights unless a representative sample from a recruitment agency or market research firm is used. Recently, Le et al. (2018) studied food allergies and found consistent results from a web-based and paper-based survey (without DCE). Similar to the trend in general DCE research, the use of online surveys through a representative panel is increasing, although face-to-face interviews with respondents who are randomly selected at stores are still in use as well. However, approximately half of these studies do not discuss the representativeness of the sample nor report the response rate. This practice has been more commonly applied in more recent studies. Irrespective of which survey mode is applied, the Suppl. Mat. includes some general guidelines that help researchers to obtain unbiased results while ensuring ethical practices.

5. Biases: origin and mitigation

While using a survey is often the only way to learn more about preferences for specific food or policy characteristics, it comes with its own challenges related to the external validity. Several biases can occur and need to be addressed in this context. The impact of biases on food consumption has been studied extensively, so there is abundant research focusing on biases in food-related DCE studies. A search in the Web of Science (see Suppl. Mat.) for publications dealing with these topics yielded more than 100 results. All of these studies were published from 2005 onwards and several of them are referred to below.

5.1. Hypothetical bias

One of the major shortcomings when using surveys to elicit preferences is hypothetical bias (e.g., Hensher, 2010) as individuals might behave inconsistently, when they do not have to back up their choices with real commitments. Respondents may not reveal their true preferences without real commitments as a DCE is not incentive compatible (Lusk & Schroeder, 2004). Thus, DCEs are said to suffer from non-consequentialism, which may lead to overestimation of WTP values and market shares and may undermine the external validity of DCEs. Meta-analyses have found that the stated WTP can be two to three times higher, on average, than the revealed WTP (e.g., List & Gallet, 2001; Little & Berrens, 2004; Murphy et al., 2005). For a recent, in-depth overview of the sources, measures, and controls of hypothetical bias in stated preference methods, we refer to Haghani et al. (2021a; 2021b).

To mitigate hypothetical bias, several approaches have been used to make DCEs more realistic and to attach consequences to choices people make. Ex-ante survey design strategies, incentive-compatible DCEs, as well as ex-post techniques, can be used to minimize or eliminate hypothetical bias (e.g., Loomis, 2014; Johnston et al., 2017; Zawojska & Czajkowski, 2017; Haghani et al., 2021a; 2021b).

An easy ex-ante option is to use cheap talk scripts, honesty oaths, or training (e.g., List & Gallet, 2001; Johnston et al., 2017). Cheap talk scripts rely on reminding respondents of the hypothetical nature of scenarios and the tendency of respondents to inflate value estimates, but they are not always effective (e.g., Carlsson et al., 2005; Murphy et al., 2005; Champ et al., 2009). Jacquemet et al. (2013) used an oath that participants signed and promised to tell the truth and provide honest answers and found that the solemn oath outperformed cheap talk in reducing hypothetical bias. Rather than using a cheap talk script, de-Magistris et al. (2013) used an implicit honesty priming task to activate honesty among primary shoppers resulting in a reduction of hypothetical bias. Recently, Drichoutis et al. (2017) used a between-sample approach to compare the impact of using no script, a cheap talk script, a consequentiality script, and a cheap talk plus consequentiality script to investigate consumers' preferences for a fair labor certificate for strawberries. They found no statistically significant effect of the scripts on the responses reflecting consumers' stated values for fair labor. Moreover, as mentioned by Johnston et al. (2017), these ex-ante methods may have implications for framing and priming and can thus introduce new biases into the results. Alternatively, visualization of alternatives in a choice set may impact the resulting WTP estimates. Dynamic visual presentation formats such as video, virtual reality (VR), or immersive virtual reality may lead to significantly lower error variance and significantly differing preference and WTP estimates compared to the traditional matrix-based textual format (Mokas et al., 2021). VR technology can offer several benefits when used in retail contexts to study purchase decisions as it can provide a more emotionally engaging customer experience and more natural user interactions such as gestures (e.g., Burke, 2018; Meissner et al., 2020). High-immersive VR shopping has potential as a tool to understand and predict consumer behavior in physical stores (e.g., Siegrist et al., 2019).

A second, generally more effective, option is to use real choice experiments[9] (RCEs), wherein the tasks are incentivized by randomly choosing one of the choice tasks as binding after the respondent has completed all of the choice tasks (for a recent overview see Haghani et al., 2021a; 2021b). The use of real products, and making participants buy the chosen product in the randomly selected binding choice task unless they select the no-buy option, makes respondents' choices more similar to real purchasing behavior (e.g., Moser et al., 2014; Liebe et al., 2019; Ballco & Gracia, 2020). For example, Chang et al. (2009) studied the ability of three preference elicitation methods (hypothetical choices, non-hypothetical choices, and non– hypothetical rankings) to predict actual retail shopping behavior in three different product categories (ground beef, wheat flour, and dishwashing liquid). Overall, they found a high level of external validity. Their results suggest that the non-hypothetical elicitation approaches, especially the non-hypothetical ranking method, outperformed the hypothetical choice experiment in predicting retail sales. Among other studies, the RCE approach has also been applied to assess preferences for beef steak (Lusk & Schroeder, 2004), salmon in Norway (Alfnes et al., 2006), canola oil in Canada (Volinskiy et al., 2009), almonds in Spain (de-Magistris and Gracia, 2014), applesauce in Italy (Bazzani et al., 2017), and yoghurt in the United States (Fang et al., 2019). Recently, RCEs have also been combined with sensory testing, which has made it possible to incorporate the effects of sensory or intrinsic attributes into the study. This enables the researcher to expand the scope of the study to repurchases rather than being limited to initial purchases. In other words, it considers search, credence, as well as experience characteristics and, as such, provides more complete and realistic information about consumer behavior in real-life. Ballco and Gracia (2020) found that after experiencing the real taste of a product, preferences change significantly compared to the

[9] See supplemental information for a note on its original inception. RCEs (aka as consequential DCE) are not to be mistaken with revealed choice modeling as the choices are still being made in an experimental setting.

initial purchase. A disadvantage of RCEs is that the products need to be available.

Another option is to actually work in a real-life setting such as a supermarket and make choices tangible. For example, Vlaeminck et al. (2014) designed an experimental food market in a natural consumer environment to investigate the impact of the new ecolabel for fruit, vegetables, and protein in Belgium. In an Australian supermarket experiment, Vanclay et al. (2011) found that sales increased by 4 percent after labeling for products with a "green light" carbon label. In such a framed field experiment, real products, and actual cash are transacted; this makes the experimental market both non-hypothetical and incentive compatible, which increases the external validity of the observed behavior (Lusk & Shogren, 2007). Recently, Wuepper et al. (2019) studied preferences for a water savings label related to coffee in a real online shop and in a hypothetical setting with cheap talk script. They found no significant preferences for the water label in the real online shop. However, the more likely respondents were to care about their appearance and the lower their self-control, the more likely they were to express a significant WTP for the water label in the hypothetical setting.

Finally, it is also possible to mitigate hypothetical bias at the data analysis phase by means of procedures that screen the data for implausible responses. This may be based on respondents' stated information about their cut-off points (Swait, 2001) – that is, minimum or maximum WTP – for the good in question (Ding et al., 2012). Alternatively, respondents can be asked how certain they are about their choice and how closely they feel it mirrors their preferences (Ready et al., 2010). Note that respondents' stated certainty may be influenced by the complexity of the choice task, learning, and fatigue (Beck et al., 2016). Results of ex-post approaches generally conclude that hypothetical bias exists and that follow-up questions can be used to improve WTP estimates, although an incorrect calibration of the responses may produce more biased results than doing nothing at all (Beck et al., 2016).

Recently, Colombo et al. (2022) compared the relative performance of ex-ante and ex-post measures that both mitigate hypothetical bias. Specifically, they tested whether ex-ante cheap talk, a reminder of the project's relative spatial extent, or a combination of both affected stated WTP. They also verified the impact on WTP estimates of an ex-post treatment wherein respondents were given the opportunity to revise choices that were identified as being inconsistent. Using a DCE on the environmental and social impacts of organic olive oil production they found that WTP estimates of treatments related to ex-ante mitigation strategies did not differ significantly from those obtained from a control treatment with standard budget constraint reminder. However, the ex-post approach resulted in a significant reduction in mean WTP estimates.

5.2. *Social desirability bias*

Individuals typically know when they take part in a research study and therefore often behave to please the researcher, avoid embarrassment, or "look good" (e.g., Costanigro et al., 2011; Norwood & Lusk, 2011). In addition, ticking the socially desirable box in a survey implies the same cost to the respondent and may give rise to a "warm glow" effect (Andreoni, 1990). In so doing, respondents misrepresent their true preferences and may systematically misreport socially sensitive behavior or attitudes (e.g., Zaller & Feldman, 1992), resulting in social desirability bias (SDB).

Several approaches have been adopted to deal with SDB: using scales, inferred valuation, or consequential valuation techniques (e.g., King & Bruner, 2000; Larson, 2019; Horiuchi et al., 2020; Haghani et al., 2021a; 2021b). Firstly, the most widely used approach to detect and control for SDB in the analysis and interpretation of the survey results are SDB scales, which are constructed by asking a series of questions designed to determine whether respondents say they engage in an activity that is socially desirable, but that is thought to rarely be acted on (e.g., Larson, 2019). An example of a scale that can be used to measure SDB is the impression management scale in the Balanced Inventory of Desirable Responding Short Form (BIDR-16, Hart et al., 2015). Secondly, rather than asking someone what choices they would make, inferred valuation entails asking what choices someone believes another person would make. To illustrate, Lusk and Norwood (2010) showed that only 16 percent of Americans agreed with the statement, "low meat prices are more important than the well-being of farm animals," while 68 percent agreed that, "the average American thinks that low meat prices are more important than the well-being of farm animals." Thirdly, making a DCE consequential (see Section 5.1) is likely to counteract SDB. Finally, it is important to note that SDB is an artifact of any study that participants are aware of, such as a survey study, and is generally absent in normal, everyday shopping experiences.

5.3. *Information bias*

It is well established in the stated preference literature that the information provision influences the responses given by survey respondents (e.g., Ajzen et al., 1996; Teisl et al., 2002; Yeh et al., 2018; Mariel et al., 2021). Essentially, an appropriate amount of information should be provided such that respondents have a clear definition of the good that they are valuing. However, providing information about a product can be viewed as persuasive communication and is likely to change the respondents' attitudes and intentions. Priming typically occurs before the respondents are asked to complete a DCE (Harris et al., 2009; Bronnmann & Asche, 2017), while framing is part of the DCE. Framing – that is, the manner in which the good or choice scenario is described – can affect the respondents' mean WTP as well as their WTP variance (e.g., Hoevenagel & van der Linden, 1993; Rousseau & Vranken, 2013; Vecchio et al., 2016; Yeh et al., 2018). (See Suppl. Mat. for more information.).

5.4. *Other biases*

Several other biases can influence consumers' intentions and behaviors (see Suppl. Mat. for additional information). Status quo bias is evident when people prefer things to stay the same by doing nothing or by sticking to a decision they made previously (Samuelson & Zeckhauser, 1988). Oehlmann et al. (2017) showed that the frequency of status quo choices is influenced by the design dimensions of the experiment (number of tasks, alternatives, attributes, levels and level range) and by the choice task complexity. The halo effect is a well-documented social-psychology phenomenon that causes people to be biased in their judgments by transferring their feelings about one attribute to other, unrelated attributes (Thorndike, 1920; Sörqvist et al., 2015; Prada et al., 2019). The country-of-origin effect (COO), also known as the nationality bias, is a psychological effect describing how consumers' attitudes, perceptions, and purchasing decisions are influenced by products' COO labeling (Nagashima, 1970; Shimp & Sharma, 1987; Yeh et al., 2018). Ethnocentrism is the term that has often been applied to the home country bias portion of the COO effect (Balabanis & Diamantopoulos, 2004). In conclusion, psychological and behavioral research has observed a plethora of consumer biases, many of which are relevant when making food-related choices.

6. Choice models and their estimation

Based on the respondents' choices, one can estimate the parameters of a discrete choice model. These parameters are often called part-worths; that is, the values that people attach to the different attribute levels. This knowledge can then be used to optimize products, to predict market shares and, if cost is among the attributes, to compute the WTP for changes in the attribute levels (e.g., Lenk et al., 1996; Green & Srinivasan, 1990; Train, 2009).

The majority of choice models have adopted a decision rule based on random utility maximization (RUM). RUM models assume that decision makers assign a utility to each alternative in the choice set and choose the alternative with the highest utility. The utility of an alternative is traditionally modeled as the sum of a linear function of the attribute levels and an error term that represents the unobserved part of utility (e.g., Train, 2009).

Estimation is typically done by maximizing the likelihood (the frequentist approach) or by simulating from the posterior distribution of the parameters (the Hierarchical Bayesian approach) (e.g., Train, 2009). The computation time of both approaches depends on a number of characteristics, such as the type of heterogeneity distribution, the number of draws, the model specification, the number of fixed and random parameters, making it impossible to predict which method is most efficient in a specific case (Train, 2001).

6.1. Multinomial logit model

The simplest model, the multinomial logit model (MNL),[10,11] includes only one set of part-worths, which can be interpreted as the average preference in the population (e.g., Alberini et al., 2006). By including interaction terms between alternative-specific and individual-specific attributes, one can capture systematic heterogeneity in these models. The parameter values that maximize the likelihood function – that is, the probability of obtaining the choices observed in the sample – are the maximum likelihood estimates. All popular software packages, such as SAS, SPSS and STATA, but also dedicated software like NLogit, and many R packages such as mlogit and multinom, can calculate the parameter estimates of the MNL model, together with their asymptotic standard errors and goodness-of-fit measures.

6.2. Random heterogeneity models

Nowadays, it is much more common to model random heterogeneity in the population. A direct, bottom-up approach that does not impose any a priori population distribution on the part-worths is the Firth penalized maximum likelihood approach (Firth, 1993; 1995) for estimating the MNL model using individual data (Kessels et al., 2019). In contrast, top-down approaches make use of distributional assumptions pooling the data from different respondents. Two broad top-down model classes are currently often used: MIXL models where the distribution of the part-worths is assumed to be continuous (also called the random parameter logit (RPL) model), and latent class (LC) models, which assume discrete part-worth values describing the different segments in the population.

6.2.1. Mixed logit models

Note that fitting a MIXL model will yield the parameters of the heterogeneity distribution (called the hyperparameters) and potentially also the estimated individual part-worths. The likelihood of MIXL models involves a multivariate integral as the probabilities have to be integrated over the heterogeneity distribution. The type of distribution has to be chosen by the user and most software packages allow users to choose from a large range of distributions. Maximizing the likelihood with respect to the hyperparameters again yields maximum likelihood estimates, but as there is no closed form expression for the multivariate integral, the maximization problem is considerably more complex (e.g., Hole, 2007).

With the simulated likelihood approach (see Suppl. Mat.), estimates are computed for the hyperparameters, but the method does not yield individual parameters. Bayes' theorem can then be used to obtain the posterior distribution of individual part-worths, conditional on the observed sequence of choices of that respondent and using the simulated likelihood estimates for the hyperparameters (see, for instance, Train, 2009).

The optimization problem in the (simulated) maximum likelihood approach can be very difficult (Train, 2009) and can give rise to convergence problems. Even if the algorithm converges, there is no guarantee that the global maximum has been obtained, and the procedure should be rerun from different starting values to check whether a better result can be found. In the Hierarchical Bayesian approach (see Suppl. Mat.), the hyperparameters and the individual part-worths are estimated simultaneously.

Computing the WTP, or more generally, computing the marginal rate of substitution based on MIXL models, can lead to computational problems. If the parameter in the denominator is small, this gives rise to numerical problems and the distribution of a ratio of distributions is not always a proper distribution. Therefore, it is recommended to estimate the model in WTP space, meaning that the model is reparametrized such that the parameters of the model are the WTP values instead of the part-worths, if WTP values are the focus of the study (see, e.g., Vermeulen et al., 2011; Scarpa et al., 2008).

Also in the food literature, as retrieved based on the query in Suppl. Mat., the MIXL model has become the established model (see, e.g., Onozaka & McFadden, 2011; Van Loo et al., 2011; 2014; Janssen & Hamm, 2012; Aprile et al., 2012; Scarpa et al., 2013), as is the case in many other disciplines. These models are mainly estimated using NLOGIT software, occasionally using STATA or Latent Gold. This explains why only results of simulated likelihood estimation can be found in this literature, as this is the only method that is implemented in these software packages[12]. Hierarchical Bayesian estimation, as well as use of the many free R packages, seems to be completely absent in the literature on DCEs for food.

6.2.2. Latent class models

LC logit models assume that the population can be divided in segments that have their own part-worths (e.g., Boxall & Adamowicz, 2002; Greene & Hensher, 2002). Estimation of the model yields the class specific part-worths, the size of each segment and the individual probabilities of belonging to each segment. The relation of the class membership to socio-demographic or other respondent-related variables can be estimated simultaneously (e.g., Jarvis et al., 2010; Nguyen et al., 2015; Rousseau, 2015).

[10] Although the terms Conditional Logit model and Multinomial Logit model are often considered to be interchangeable, strictly speaking they are not. Multinomial models have alternative-specific parameters and use the characteristics of the decision maker to explain choice behavior. Conditional logit models use generic parameters to explain choice behavior by the characteristics of the alternatives. Because generic parameters are independent of the alternative, variables that have the same value for all the alternatives in a choice set (such as individual-specific variables) cannot be included as such in a conditional logit model because they would drop from the likelihood function. To include such variables, one must include interaction terms between these variables and the attribute levels (see, for instance, Hoffman & Duncan, 1988).

[11] Logit models that assume Extreme Value Type I error distributions for the utility are much easier to handle than probit models that assume normally distributed errors. Logit models have closed-form expressions for the probabilities and, as a result, they are much easier to interpret and to use than probit models (Train, 2009). Moreover, McFadden and Train (2000) proved that the MIXL model (that is, MNL with random parameters) can approximate any choice model to any degree of accuracy with appropriate choice of variables and mixing distribution. Therefore, a MIXL model can approximate any multinomial probit model, but the reverse is not true.

[12] Without being exhaustive, we list some software packages for the models and methods described in Section 7.2: the R packages mlogit, gmnl, Apollo, and ChoiceModelR can be used for simulated maximum likelihood estimation of MIXL models (among many other models); the R packages bayesm, RSGHB, and Apollo for Hierarchical Bayesian estimation of (among other models) MIXL models; and the R packages gmnl and BayesLCA estimate LC models with simulated maximum likelihood and Hierarchical Bayes estimation, respectively.

Also for LC models, maximum likelihood estimation and Bayesian estimation have been implemented, as well as the expectation–maximization method, which iteratively computes the class memberships given the parameters, and maximizes the likelihood with respect to the parameters, given the class memberships (e.g., Train, 2008). These methods can also be used in combination with each other; that is, starting with a few iterations of the EM algorithm and using these results as starting values for the maximum likelihood approach (e.g., Meulders, 2013; Vermunt & Magidson, 2005). This can be combined with some priors on the parameters to prevent boundary solutions. The number of classes has to be specified by the user who typically fits models with different number of classes and then selects the appropriate number based on fit statistics, which have a penalty for the complexity of the model such as Akaike's information criterion (AIC) or the Bayesian information criterion (BIC) and variants thereof.

LC models can suffer from identification problems (Vermunt, 2003), meaning that several parameter values yield the same likelihood value. Running the algorithm from different starting values is a simple method to detect the problem. On the other hand, weak identification means that the data is not informative enough to obtain stable results. This is apparent from large standard errors and/or slow convergence.

Systematic heterogeneity, modelled by interactions between alternative-specific and individual-specific attributes, and random heterogeneity, which is modeled by random parameters, have also been combined (e.g., Asioli et al., 2016a). Furthermore, individual level partworths resulting from a mixed logit model, as well as class-level partworths resulting from a latent class model, have subsequently been investigated to detect systematic differences by regression, PCA, cluster analysis and other statistical methods (e.g., Asioli et al. 2016b, 2018; Greene & Hensher, 2003). Collecting such information is important to explain (part of) participants' preference heterogeneity (Bechtold & Abdulai, 2014). To this end, hybrid choice models with latent variables measuring consumers' attitudes are also increasingly used (e.g., Walker & Ben-Akiva, 2002; Mariel et al., 2021). For example, Palma et al. (2018) compare three approaches to consider preference heterogeneity in a DCE: (i) systematic preference variations based on sociodemographic characteristics; (ii) latent classes; and (iii) hybrid choice models with latent variables measuring consumers' attitudes. Based on an example measuring wine preferences in Chile, they conclude that the most appropriate approach depends on the research objectives. Additionally, to account for the correlation between alternatives, the mixed logit with an error component may be used, as proposed by Scarpa et al. (2005), and put to use in food DCEs, as done, for example, in Caputo et al. (2013) and Scarpa et al. (2013).

7. Random regret models as an alternative to random utility models

In the past decade, discrete choice models based on random regret minimization (RRM) have been introduced as an alternative to random utility models (Chorus et al., 2008; Chorus, 2010). Instead of the common assumption that respondents maximize their utility, RRM models assume that decision makers try to minimize so-called anticipated regret (e.g., Zeelenberg & Pieters, 2007) which emerges if the considered alternative is outperformed by one or more competing alternatives on some attribute level(s) (e.g., Chorus et al., 2008). The systematic regret associated with an alternative is then obtained by summing the attribute level regrets generated by all competing alternatives across all attributes (e.g., Chorus et al., 2008; Chorus, 2010, 2012). Similar to RUM models, RRM models assume that the total random regret associated with an alternative is the sum of a systematic component and a random error term. Assuming an Extreme Value Type-I distribution for the negative of the errors, RRM models also feature a Multinomial Logit formulation for the choice probabilities (e.g., Chorus, 2010, 2012). For more information, see Suppl. Mat.

In an overview of the empirical RRM literature, Chorus et al. (2014) concluded that the RRM decision framework performs better when explaining choices that are considered difficult or important, or when the choice outcome will also be evaluated by others. As many food decisions are habitual (e.g., Adamovicz & Swait, 2012; van't Riet et al., 2011), this helps explain why the RUM model is still the dominant model used when explaining decision-making in the food domain. However, regret-based models can be more appropriate when food decisions are important and/or will be evaluated by others - such as buying food for a special occasion (e.g., Biondi et al., 2019) - or when food safety is an issue - such as making a food decision after a food recall (e.g., Dennis et al., 2020).

7.1. Comparison of RUM and RRM approaches

A comparison of papers that apply both RUM and RRM models (Chorus et al., 2014) shows that differences in model fit and out-of-sample performance are usually small, and that neither of the models can generally be regarded as superior. However, there is some evidence that RRM models may be more appropriate for choices that are regarded as difficult to make, or important, or in which one needs to justify the choice made to others. For instance, Wang et al. (2017) and Wang et al. (2018b) found that RRM models fit much better than RUM models on choices in an emergency context. Furthermore, Hess et al. (2014) showed that the inclusion of a "none of these" opt-out has a detrimental effect on the fit of RRM models (but not on RUM), whereas the opposite holds when an "I am indifferent" opt-out is included. Moreover, differences in model fit between RUM and RRM may also be more pronounced when the model includes a more complex model specification (such as nested logit or mixed logit). Although differences in fit between RUM and RRM are often small, the predictions made by both types of models can differ substantially, leading to different market share forecasts, different attribute elasticities (e.g., Hensher et al., 2013), and different managerial or policy implications (e.g., Chorus et al., 2014; van Cranenburgh & Chorus, 2018).

Finally, only two papers were found that used the RRM model to model consumer choice in the food domain. Biondi et al. (2019) observed that food choice decisions are sometimes perceived as more difficult or important and look into the question of what framework governs the decision-making process. Their conclusions confirm the results found in other domains, that the RRM model returns coherent estimates of anticipated regret and is not inferior to the RUM model in terms of goodness of fit and prediction. Dennis et al. (2020) used a DCE to study consumers' food decisions when purchasing beef after a food recall. Using an LC model, they identified 40 percent of the consumers as utility maximizers and 60 percent as regret minimizers, indicating that using RRM is more appropriate for modelling risky choices. They also found substantially different price discounts of a food recall using RRM and RUM, which shows the importance of selecting an appropriate decision rule.

8. Heuristics

We typically assume that respondents assess and make a trade-off between all attributes describing the alternatives in a fully compensatory way (e.g., Chorus, 2014). This assumes that respondents attend to the complete set of information and consider all attributes/attribute levels and all alternatives presented when choosing their preferred alternative. However, respondents may not process all information presented in a rational way. This can occur due to the cognitive burden and the task complexity, especially when facing a large amount of information in choice tasks with many attributes and attribute levels (Simon, 1955; Payne, 1976). As a result, respondents may not behave fully rationally, and instead use simplifying decision rules, also named heuristics, to reduce cognitive effort and to help make choices (Shah & Oppenheimer, 2008). There is growing empirical evidence that choice behavior is not always fully compensatory, so we can no longer assume that all attributes, attribute levels, and alternatives are fully processed (e.g. Hensher et al., 2005; Campbell et al., 2011). Two popular heuristics in the CE literature are attribute-non-attendance (ANA, Hensher et al., 2005; Kragt, 2013) and consideration-set screening (CSS, e.g. Hauser, 2014).

8.1. Attribute non-attendance

ANA refers to the lexicographic decision heuristic in which respondents ignore some of the attributes in a choice task. Hensher (2006, 2014) mentioned that not only does task complexity induce ANA, but also the relevance of the information counts, with less relevant attributes being more likely to be ignored (relevance simplification rule). However, Weller et al. (2014) reported only a weak relation between ANA and design dimensions such as number of attributes and attributes level. Alemu et al. (2013) identified behavioral reasons that cause ANA, such as low preference for or importance of the attribute or a disinterest in the attribute, design-related issues such as the complexity and cognitive demand associated with the choice task, as well as unrealistic attribute levels. Empirical evidence demonstrates that accounting for ANA has implications for key outputs such as the marginal WTP estimates (e.g., Hensher et al., 2005; Hensher, 2006), predicted probabilities, and market share predictions (e.g., Scarpa et al., 2013).

There are two general approaches to examine and account for ANA in DCE (Hensher, 2014). Firstly, stated ANA relies on self-reported attendance by asking respondents follow-up questions on attributes they have ignored, either after each choice task (choice task stated ANA) (Puckett & Hensher, 2008; Scarpa et al., 2010) or after the whole sequence of choice tasks (serial stated ANA) (Hensher et al., 2005; Alemu et al., 2013). Secondly, ANA can be inferred ex-post based on the observed choices (inferred ANA) (Caputo et al., 2018; Scarpa et al., 2013). Caputo et al. (2018) and Hess and Hensher (2010) reported that stated and inferred attribute processing are not always consistent, so both approaches may be complementary. For more information, see Suppl. Mat. In addition to ANA, which assumes that certain attributes are ignored, Erdem et al. (2015) suggested accounting for attribute-level non-attendance as their empirical evidence shows that attribute processing differs across the attribute levels.

Most of the research contributions on ANA in DCE come from the fields of transportation (e.g., Hensher et al., 2005), health economics (e.g., Erdem et al., 2015), and environmental economics (e.g., Campbell et al., 2011). However, more recently, ANA has been studied in the field of food economics (e.g., Caputo et al., 2018; Scarpa et al. 2013). When evaluating the main journals publishing food-related DCEs as retrieved with the queries specified in Suppl. Mat., over 30 articles were identified that cover the topic of ANA. Most publications dealt with inferred ANA often combined with stated ANA, while only a small fraction addressed only stated ANA. When evaluating the type of inferred methods applied, the use of equality constrained latent class (ECLC) models is the most common inferred method, followed by the Hess and Hensher method. Scarpa et al. (2013) concluded that the use of ECLC better aligns with the stated ANA as compared to the Hess and Hensher approach. Future research may investigate strategies to reduce ANA in the context of food choices. Bello and Abdulai (2016) found that honesty priming can help reduce the rate of ANA among respondents. Some recent studies on food choice have shown how ECLC can be used to infer the probability that respondents belong to a latent class where choices are made randomly (Caputo et al., 2018). Therefore, this class consists of inattentive respondents (Malone & Lusk, 2018) and is a novel way to deal with data quality problems (measurement error) caused by respondents' inattentiveness and random choices.

8.2. Consideration-set screening

CSS implies that respondents only consider part of the alternatives in the choice set when making a choice (Payne, 1976). That is, when faced with many alternatives to choose from, respondents may resort to a "consider then choose" decision process. In a first screening stage, respondents may use heuristics (e.g., Hauser, 2014) or apply relevant constraints (e.g., Swait & Ben-Akiva, 1987) to identify a smaller consideration-set of (feasible) alternatives that need further evaluation, and in a second stage they may adopt a standard compensatory model to choose from the consideration-set (e.g., Bettman et al., 1998). Research has shown that accounting for CSS often leads to models that fit the data better, have better predictive power, and provide a more realistic description of the choice process (e.g., Chorus, 2014; Leong & Hensher, 2012).

Many heuristic decision rules for CSS have been described in the literature, especially in marketing and transportation (for a review, see Hauser, 2014, Leong & Hensher, 2012). For instance, elimination by aspects (e.g., Gilbride & Allenby, 2006; Erdem et al., 2014) means that respondents eliminate alternatives with unacceptable attribute levels until one alternative remains. Satisficing (Simon, 1955, González-Valdés & Ortúzar, 2018) implies that respondents evaluate alternatives sequentially and choose the first alternative that has an acceptable utility level. Lexicographic choice (e.g., Jedidi & Kohli, 2008; Kohli & Jedidi, 2007) means that respondents systematically select the alternative that scores best on one attribute and ignore all other attributes (for example, always choosing the product with the lowest price). Conjunctive screening occurs when respondents only consider alternatives for which all attributes have an acceptable level, whereas disjunctive screening implies that alternatives are in the consideration set if at least one attribute has an acceptable level (e.g., Gilbride & Allenby, 2004; Cantillo & Ortúzar, 2005). Subset-conjunctive screening means that an alternative is considered if at least k (out of m) attributes have an acceptable level (Jedidi & Kohli, 2005; Kohli & Jedidi, 2005). Finally, using disjunctions of conjunctions (Hauser et al., 2010), an alternative is in the consideration set if at least one of multiple conjunctive criteria is satisfied (for example, either Attributes A and B have acceptable levels or Attributes C and D have acceptable levels).

8.3. Integration of CSS heuristics into choice models

Several modeling strategies have been used to integrate CSS heuristics into choice models with the two-stage model of Manski (1977) being an especially popular approach. As the assumption that all participants use the same type of heuristic decision rule is increasingly seen as being unrealistic, more recent research has focused on modelling heterogenous decision rules (e.g. Adamowics & Swait, 2012; González-Valdés & Raveau, 2018). More information on this topic is provided in the Suppl. Mat.

In sum, several papers have indicated that models including heuristics may provide a more realistic analysis of food choices, with more accurate choice predictions and welfare estimates (e.g. Scheibehenne et al., 2007; Sawada et al. 2014; Peschel et al., 2016; Sandorf & Campbell, 2019). However, a search in Web of Science indicates that most papers still rely on standard MNL or MIXL models. A Web of Science search as specified in the Suppl. Mat. in food journals or food-related journals for papers that include a choice-experiment yielded only ten papers that also include keywords related to choice heuristics. Consequently, developing new models to understand heuristics that affect food choice, and investigating how individual differences in decision making affect food choices, remain important directions of future research in sensory and consumer science (Jaeger et al., 2017).

9. Reliability and validity of food-related DCEs

Several criteria can be used to assess the quality of DCEs and their outcomes. Researchers often aim to make correct inferences, both about what is actually studied (internal validity) and about what the results generalize to (external validity) (Persson & Wallin, 2015). Recently, Bishop and Boyle (2019) discussed reliability as well as three aspects of validity – content validity, construct validity, and criterion validity – as criteria for considering the accuracy of value estimates obtained from non-market valuation surveys.

As a first criterion, reliability describes the extent to which a particular test, such as a survey, will produce similar results in different circumstances assuming nothing else has changed (Roberts & Priest, 2006). Typical valuation studies such as DCEs only involve one measurement point, such as a single survey, so nothing can be said about their reliability as a method to derive estimates of welfare change related to food choices. Test–retest studies are the main tool used to assess the reliability of survey-based measurements (Liebe et al., 2012; Mørkbak & Olsen, 2015; Foerde et al., 2018). Participants are asked to complete the same DCE at more than one point in time and hence provide independent observations. This retesting can be done with the same subjects (within-subject test–retest) or with a different sample from the same population (between-subject test–retest) (Zeller & Carmines, 1980). An interesting discussion and application of a test–retest study for consumer preferences for beef produced via traditional or innovative production processes can be found in Rigby et al. (2016). Besides testing the quality of the research, test–retesting can also be informative when studying repeat purchase decisions, as Williamson et al. (2016, 2017) did for wine (repeat) purchases in China. They used the test–retest setting to disentangle the effects of search characteristics such as country of origin and experience characteristics such as taste.

As a second set of criteria, validity focuses on the closeness of what we believe we are measuring to what we intended to measure (Roberts & Priest, 2006). Following Zeller and Carmines (1980), Bishop and Boyle (2019) distinguished three subcategories when looking into the validity of valuation methods. Firstly, content validity focuses on the extent to which the different components and procedural steps of a DCE survey allow the researchers to measure the true preferences (Bishop & Boyle, 2019). Secondly, construct validity focuses on the value estimates and how the validity of these might be assessed in the absence of knowledge about the true values (Bishop & Boyle, 2019). A key element of construct validity is the so-called expectation-based validity (Mariel et al., 2021). An analyst will often have some prior expectations of the values and how they relate to other variables. Sources of such expectations can be economic theory, intuition, or past empirical evidence. For example, Mariel et al. (2021) noted that, based on the economic law of demand, the most crucial validity test that any DCE survey has to pass is that increasing the cost of an alternative should decrease the probability of choosing that alternative, keeping everything else constant. Thirdly, criterion validity involves comparing results from two valuation methods (Bishop & Boyle, 2019). For example, comparing the WTP estimates obtained in a new DCE survey to previously obtained highly valid WTP estimates for the same good such as market prices, simulated markets or incentive-compatible field or lab experiments (Mariel et al., 2021).

Clearly, many of the factors presented in the previous sections have an impact on the reliability and validity of food-related DCEs. However, this topic has not yet been studied extensively.[13] An interesting method to increase one's insights into the reliability and validity of DCE studies is the *meta*-analysis. A *meta*-analysis is a statistical method used to combine results from the relevant studies, and the resulting larger sample size can provide greater reliability (precision) of the estimates of any treatment effect (Møller & Myles, 2016). The use of a *meta*-analysis can be an interesting tool to assess the quality of DCEs. For example, the *meta*-analyses of Little and Berrens (2004) and Murphy et al. (2005) allowed us to learn more about the hypothetical bias of stated preference methods. The value of a *meta*-analysis depends heavily on the quantity, quality, and heterogeneity of the included studies, as well as a clear and detailed methodology. The PRISMA guidelines provide an overview of all essential elements of systematic reviews and *meta*-analyses (Moher et al., 2009). We were only able to find a limited number of *meta*-analyses related to food choice and stated choices. We could only find three studies published before 2015: Lusk et al. (2005) and Dannenberg (2009) studied genetically modified food valuation studies, while Tully and Winer (2014) investigated the role of the beneficiary in estimating WTP values of socially responsible products (including food products). More recent *meta*-analyses have dealt with health claims (Kaur et al., 2017; Dolgopolova & Teuber, 2018), biofortified foods (De Steur et al., 2016), credence characteristics of livestock products (Yang, & Renwick, 2019), food safety in China (Yang & Fang, 2020), local food production (Printezis et al., 2019) and sustainable food (Bastounis et al., 2021; Li & Kallas, 2021).

10. Conclusion and recommendations

Discrete choice experiments (DCEs) have been the most used valuation method to uncover preferences and elicit willingness to pay (WTP) or willingness to accept (WTA) for foodstuffs, in particular for meat, organic foods, and functional foods or foods with nutrition or health claims. Unlike other valuation methods, a DCE mimics the architecture of a consumer's buying decision taking place in stores, shops or restaurants as it allows a comparison across alternatives. Yet, DCE methodology is not limited to food-related preferences or food choice. Hence, we set out to compare common practice in food research to general DCE best-practice in order to promote knowledge discovery, particularly for junior researchers and scholars, and knowledge creation through recombination. We find that over recent decades, the use of DCEs in food has covered a wide breadth of applications and has shown significant methodological progress. Yet, several useful methodological innovations did not get adopted equally across research fields. Below, we highlight those innovations that are particularly useful for food DCEs in view of increasing their reliability and validity.

Food DCEs have tended to focus on familiar food products and their credence attributes. More recently, unfamiliar food products, such as novel proteins, insect-based food products or cultured meat have come to the forefront. This evolution calls for increased attention to the reporting of the findings of a qualitative exploratory phase as this may improve the DCE's content validity. The process of attribute selection in the food DCE literature was found to still have scope for improvement as the development process of attributes and attribute levels is often given little attention and tends to be poorly documented. Yet, the number of attributes and attribute levels has a significant impact on the ease of understanding the choice question by study participants. Investigating insights from previous studies, organizing expert interviews and focus group discussions will provide key information on the participants' decision process in a given context. A qualitative pre-phase ensures the relevance of what is being measured, helps in targeting the right respondents and choice context, and feeds the experimental design the design dimensions it needs to compute the statistical efficiency while taking respondent efficiency into account.

With regards to computing the statistical efficiency of an experimental design, the Bayesian D-optimal design approach is increasingly being implemented and considered state of the art for food-related DCEs. Consequently, the field is moving away from orthogonal designs, which are most suited for estimating linear models. Yet, there is still room for further methodological innovation, e.g. concerning the adoption of designs involving a no-choice option.

Compared to other fields, food DCEs have the advantage of being more easily amenable to settings that resemble the actual decision-making environment such as a virtual, mock online or actual store. They owe this feature to the fact that foodstuffs are actually being sold on markets and in (online) stores, unlike human health and

[13] Carlsson et al. (2005) investigated the impact of cheap talk scripts on the hypothetical bias (see Section 5.2.1) associated with DCEs measuring preferences for chicken and ground beef in Sweden. More recently, Lagerkvist et al. (2014) studied the reliability and validity of DCEs to measure the impact of country of origin on beef consumption decisions in Sweden using the R-index to detect transitivity and dominance in choices.

environmental quality, for example. This may explain why we found that applications of non-hypothetical or real choice experiments (RCEs) have almost exclusively emerged in DCEs targeting food choice. This is not to say that such an application is straightforward as the researcher is still bound to abiding ethical considerations with regards to the claims that are explicitly being made regarding the attributes of the goods under study. Moreover, to avoid deception, the good needs to be available at the time of study. As an additional plus, RCEs are compatible with sensory (taste, smell, appearance) testing which provides more complete and realistic information about consumer behavior in real-life. Additionally, randomly choosing one of the choice tasks as binding after the respondent has completed all of the choice tasks is expected to also reduce social desirability bias.

Finally, the goal of a food valuation study is to estimate preferences and their derived outcomes such as the willingness to pay (WTP). Inspired by the qualitative phase or pilot study the researcher may already have developed a sense of the decision rule(s) and heuristics respondents use while choosing. Model fit may provide further evidence of the decision rule that was used after estimation. We find food research to have adopted random utility based mixed (multinomial) logit models (MIXL) estimated by maximizing the simulated likelihood as the state of the art. Yet, Hierarchical Bayesian estimation can be used to avoid local optima and convergence problems that are inherent to simulated maximum likelihood estimation. To date, this analytical approach is rarely implemented in food choice analysis despite (free) software being available. Moreover, regret-minimization based models are also rarely estimated despite estimation routines being available, although such models may provide a better description of food choices than utility-maximization based models when food decisions are important and/or will be evaluated by others, or when food safety is an issue. In turn, being freely available cannot be said for software (packages) that incorporates heuristics other than attribute non-attendance, whereas several food choice papers have shown such estimation to result in more accurate choice predictions and welfare estimates.

CRediT authorship contribution statement

Sebastien Lizin: Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Sandra Rousseau:** Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Roselinde Kessels:** Investigation, Writing – original draft, Writing – review & editing. **Michel Meulders:** Investigation, Writing – original draft, Writing – review & editing. **Guido Pepermans:** Investigation, Writing – original draft, Writing – review & editing. **Stijn Speelman:** Investigation, Writing – original draft, Writing – review & editing. **Martina Vandebroek:** Investigation, Writing – original draft, Writing – review & editing. **Goedele Van Den Broeck:** Investigation, Writing – original draft, Writing – review & editing. **Ellen J. Van Loo:** Investigation, Writing – original draft, Writing – review & editing. **Wim Verbeke:** Investigation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the support of the research foundation Flanders (FWO) that enabled the creation of the scientific research network (WOG Empirische en methodologische uitdagingen in keuze-experimenten W001517N) that led to this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodqual.2022.104678.

References

Adamovicz, W., & Swait, J. D. (2012). Are food choices really habitual? Integrating habits, variety-seeking and compensatory choice in a utility-maximizing framework. *American Journal of Agricultural Economics, 95*(1), 17–41.

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2020). Best-practice recommendations for producers, evaluators, and users of methodological literature reviews. *Organizational Research Methods, 1094428120943281*.

Ajzen, I., Brown, T. C., & Rosenthal, L. H. (1996). Information bias in contingent valuation: Effects of personal relevance, quality of information, and motivational orientation. *Journal of Environmental Economics and Management, 30*(1), 43–57.

Alberini, A., Longo, A., & Veronesi, M. (2006). Basic statistical models for stated choice studies. In B. J. Kanninen (Ed.), *Valuing environmental amenities using stated choice studies* (pp. 203–227). Dordrecht: Springer.

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and Resource Economics, 54*(3), 333–359.

Alemu, M. H., Olsen, S. B., Vedel, S. E., Pambo, K. O., & Owino, V. O. (2017). Combining product attributes with recommendation and shopping location attributes to assess consumer preferences for insect-based food products. *Food Quality and Preference, 55*, 45–57.

Alfnes, F., Guttormsen, A. G., Steine, G., & Kolstad, K. (2006). Consumers' willingness to pay for the color of salmon: A choice experiment with real economic incentives. *American Journal of Agricultural Economics, 88*(4), 1050–1061.

Alphonce, R., & Alfnes, F. (2017). Eliciting consumer WTP for food characteristics in a developing context: Application of four valuation methods in an African market. *Journal of Agricultural Economics, 68*(1), 123–142.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal, 100*(401), 464–477.

Angeliki, N. M., Olsen, S. B., & Tsagarakis, K. P. (2016). Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *Journal of Choice Modelling, 18*, 18–50.

Aprile, M., Caputo, V., & Nayga, R. (2012). Consumers' valuation of food quality labels: The case of the European geographic indication and organic farming labels. *International Journal of Consumer Studies, 36*, 158–165.

Asioli, D., Almli, V., & Næs, T. (2016). Comparison of two different strategies for investigating individual differences among consumers in choice experiments. A case study based on preferences for iced coffee in Norway. *Food Quality and Preference, 54*, 79–89.

Asioli, D., Næs, T., Øvrum, A., & Almli, V. L. (2016). Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway. *Food Quality and Preference, 48*, 174–184.

Asioli, D., Berget, I., & Naes, T. (2018). Comparison of Different Clustering Methods for investigating Individual Differences using Choice Experiments. *Food Research International, 111*, 371–378.

Asioli, D., Mignani, A., & Alfnes, F. (2021). Quick and easy? Respondent evaluations of the Becker–DeGroot–Marschak and multiple price list valuation mechanisms. *Agribusiness, 37*(2), 215–234.

Balabanis, G., & Diamantopoulos, A. (2004). Domestic country bias, country-of-origin effects, and consumer ethnocentrism: A multidimensional unfolding approach. *Journal of the Academy of Marketing Science, 32*(1), 80–95.

Ballco, P., & Gracia, A. (2020). An extended approach combining sensory and real choice experiments to examine new product attributes. *Food Quality and Preference, 80*, Article 103830.

Bastounis, A., Buckell, J., Hartmann-Boyce, J., Cook, B., King, S., Potter, C., … Jebb, S. A. (2021). The Impact of Environmental Sustainability Labels on Willingness-to-Pay for Foods: A Systematic Review and Meta-Analysis of Discrete Choice Experiments. *Nutrients, 13*(8), 2677.

Bateman, I. J., Carson, R. T., Day, B., Hanemann, W. M., Hanley, N., Hett, T., JonesLee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R., & Swanson, J. (2002). *Economic valuation with stated preferences techniques: A manual*. Cheltenham, UK: Edward Elgar.

Bazzani, C., Caputo, V., Nayga, R. M., Jr, & Canavari, M. (2017). Revisiting consumers' valuation for local versus organic food using a non-hypothetical choice experiment: Does personality matter? *Food Quality and Preference, 62*, 144–154.

Becerra, M., & Goos, P. (2021). Bayesian I-optimal designs for choice experiments with mixtures. *Chemometrics and Intelligent Laboratory Systems, 217*, Article 104395.

Bechtold, K. B., & Abdulai, A. (2014). Combining attitudinal statements with choice experiments to analyze preference heterogeneity for functional dairy products. *Food Policy, 47*, 97–106.

Beck, M. J., Fifer, S., & Rose, J. M. (2016). Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. *Transportation Research Part B: Methodological, 89*(C), 149–167.

Bello, M., & Abdulai, A. (2016). Impact of ex-ante hypothetical bias mitigation methods on attribute non-attendance in choice experiments. *American Journal of Agricultural Economics, 98*(5), 1486–1506.

Bennett, J. W., & Birol, E. (Eds.). (2010). *Choice Experiments in Developing Countries: Implementation, Challenges and Policy Implications*. Cheltenham, UK: Edward-Elgar Publishing.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). *Journal of Consumer Research, 25*(3), 187–217.

Biondi, B., Van der Lans, I., Mazzocchi, M., Fischer, A. R. H., Van Trijp, H. C. M., & Camanzi, L. (2019). Modelling consumer choice through the random regret minimization model: An Application in the food domain. *Food Quality and Preference, 73*, 97–109.

Bishop, R. C., & Boyle, K. J. (2019). Reliability and validity in nonmarket valuation. *Environmental and Resource Economics, 72*(2), 559–582.

Bliemer, M. C. J., & Collins, A. T. (2016). On determining priors for the generation of efficient stated choice experimental designs. *Journal of Choice Modelling, 21*, 10–14.

Bliemer, M. C. J., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological, 44*(6), 720–734.

Boxall, P. C., & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach. *Environmental and Resource Economics, 23*(4), 421–446.

Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R., & Rose, J. (2016). Investigating internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. *Environmental and Resource Economics, 64*(3), 401–419.

Brace, I. (2018). *Questionnaire design: How to plan, structure and write survey material for effective market research*. Kogan Page Publishers.

Bronnmann, J., & Asche, F. (2017). Sustainable seafood from aquaculture and wild fisheries: Insights from a discrete choice experiment in Germany. *Ecological Economics, 142*, 113–119.

Bublitz, M. G., Peracchio, L. A., & Block, L. G. (2010). Why did I eat that? Perspectives on food decision making and dietary restraint. *Journal of Consumer Psychology, 20*(3), 239–258.

Burke, R. R. (2018). Virtual reality for marketing research. In *Innovative research methodologies in management* (pp. 63–82). Cham: Palgrave Macmillan.

Campbell, D., Hensher, D. A., & Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: A latent class specification. *Journal of Environmental Planning and Management, 54*, 1061–1076.

Canavari, M., Drichoutis, A. C., Lusk, J. L., & Nayga, R. M., Jr (2019). How to run an experimental auction: A review of recent advances. *European Review of Agricultural Economics, 46*(5), 862–922.

Cantillo, V., & Ortúzar, J. de D. (2005). A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B, 39*, 641–657.

Caputo, V., Nayga, R. M., & Scarpa, R. (2013). Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics, 57*(4), 465–482.

Caputo, V., Van Loo, E. J., Scarpa, R., Nayga, R. M., & Verbeke, W. (2018). Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments. *Journal of Agricultural Economics, 69*(1), 35–57.

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2005). Using cheap talk as a test of validity in choice experiments. *Economics Letters, 89*(2), 147–152.

Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation research part B: Methodological, 39*(7), 621–640.

Champ, P. A., Moore, R., & Bishop, R. C. (2009). A comparison of approaches to mitigate hypothetical bias. *Agricultural and Resource Economics Review, 38*(2), 166–180.

Chang, J. B., Lusk, J. L., & Norwood, F. B. (2009). How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics, 91*(2), 518–534.

Chorus, C., van Cranenburgh, S., & Dekker, T. (2014). Random Regret minimization for consumer choice modeling: Assessment of empirical evidence. *Journal of Business Research, 67*, 2428–2436.

Chorus, C. G. (2010). A new model of random regret minimization. *European Journal of Transport and Infrastructure Research, 10*(2), 181–196.

Chorus, C. (2012). Random regret minimization: An overview of model properties and empirical evidence. *Transport Reviews: A Transnational Transdisciplinary Journal, 32*(1), 75–92.

Chorus, C. G. (2014). Capturing alternative decision rules in travel choice models: A critical discussion. In S. Hess, & A. Daly (Eds.), *Handbook of Choice Modelling* (pp. 290–310). Gheltenham: Edward Elgar Publishing Limited.

Chorus, C. G., Arentze, T. A., & Timmermans, H. J. P. (2008). A random regret-minimization model of travel choice. *Journal of Transportation Research Part B: Methodological, 42*, 1–18.

Czajkowski, M., Hanley, N., & LaRiviere, J. (2015). The effects of experience on preferences: Theory and empirics for environmental public goods. *American Journal of Agricultural Economics, 97*(1), 333–351.

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: Issues and recommendations. *Health Economics, 21*(6), 730–741.

Colombo, S., Budziński, W. Czajkowski, M. & Glenk, K. (2022). The relative performance of ex-ante and ex-post measures to mitigate hypothetical and strategic bias in a stated preference study. Journal of Agricultural Economics, 1-29, doi.org/10.1111/1477-9552.12484.

Corrigan, J. R., Depositario, D. P. T., Nayga, R. M., Jr, Wu, X., & Laude, T. P. (2009). Comparing open-ended choice experiments and experimental auctions: An application to golden rice. *American Journal of Agricultural Economics, 91*(3), 837–853.

Costanigro, M., McFadden, D. T., Kroll, S., & Nurse, G. (2011). An in-store valuation of local and organic apples: The role of social desirability. *Agribusiness, 27*(4), 465–477.

Czine, P., Török, Á., Pető, K., Horváth, P., & Balogh, P. (2020). The impact of the food labeling and other factors on consumer preferences using discrete choice modeling—The example of traditional pork sausage. *Nutrients, 12*, 1768.

Dannenberg, A. (2009). The dispersion and development of consumer preferences for genetically modified food–A meta-analysis. *Ecological Economics, 68*(8–9), 2182–2192.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics, 16*(1), 67–88.

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample size requirements for discrete-choice experiments in healthcare: A practical guide. *The Patient-Patient-Centered Outcomes Research, 8*(5), 373–384.

De Marchi, E., Caputo, V., Nayga, R. M., Jr, & Banterle, A. (2016). Time preferences and food choices: Evidence from a choice experiment. *Food Policy, 62*, 99–109.

De Steur, H., Wesana, J., Blancquaert, D., Van Der Straeten, D., & Gellynck, X. (2016). Methods matter: A meta-regression on the determinants of willingness-to-pay studies on biofortified foods. *Annals of the New York Academy of Sciences, 1290*(1), 34–46.

de-Magistris, T., & Gracia, A. (2014). Do consumers care about organic and distance labels? An empirical analysis in Spain. *International Journal of Consumer Studies, 38*(6), 660–669.

de-Magistris, T., Gracia, A., & Nayga, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics, 95*(5), 1136–1154.

Dennis, E. J., Ajewole, K., Bergtold, J. S., & Schroeder, T. C. (2020). Consumer reactions to E. coli and Antibiotic residue recalls: Utility Maximization vs. regret minimization. Frontiers in Veterinary. *Science, 7*, 611.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management, 44*(1), 123–143.

Dieterle, J. M. (2022). Agency and autonomy in food choice: Can we really vote with our forks? *Journal of Agricultural & Environmental Ethics, 35*(1), 5.

Ding, Y., Veeman, M. M., & Adamowicz, W. L. (2012). The influence of attribute cutoffs on consumers' choices of a functional food. *European Review of Agricultural Economics, 39*(5), 745–769.

Dolgopolova, I., & Teuber, R. (2018). Consumers' willingness to pay for health benefits in food products: A meta-analysis. *Applied Economic Perspectives and Policy, 40*(2), 333–352.

Drichoutis, A. C., Vassilopoulos, A., Lusk, J. L., & Nayga, R. M. (2017). Consumer preferences for fair labour certification. *European Review of Agricultural Economics, 44*(3), 455–474.

Dupont, W. D., & Plummer, W. D. (1990). Power and sample size calculations: A review and computer program. *Controlled Clinical Trials, 11*(2), 116–128.

Enneking, U. (2004). Willingness-to-pay for safety improvements in the German meat sector: The case of the Q&S label. *European Review of Agricultural Economics, 31*(2), 205–223.

Enneking, U., Neumann, C., & Henneberg, S. (2007). How important intrinsic and extrinsic product attributes affect purchase decision. *Food Quality and Preference, 18*(1), 133–138.

Erdem, S., Campbell, D., & Thompson, C. (2014). Elimination and selection by aspects in health choice experiments: Prioritising health service innovations. *Journal of Health Economics, 38*, 10–22.

Erdem, S., Campbell, D., & Hole, A. R. (2015). Accounting for attribute-level non-attendance in a health choice experiment: Does it matter? *Health Economics, 24*(7), 773–789.

Fang, D., Nayga, R. M., Snell, H. A., West, G. H., & Bazzani, C. (2019). Evaluating USA's new nutrition and supplement facts label: Evidence from a non-hypothetical choice experiment. *Journal of Consumer Policy, 42*(4), 545–562.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*, 27–38.

Firth, D. (1995). Amendments and corrections: Bias reduction of maximum likelihood estimates. *Biometrika, 82*, 667.

Foerde, K., Gianini, L., Wang, Y., Wu, P., Shohamy, D., Walsh, B. T., & Steinglass, J. E. (2018). Assessment of test-retest reliability of a food choice task among healthy individuals. *Appetite, 123*, 352–356.

Gelaw, F., Speelman, S., & Van Huylenbroeck, G. (2016). Farmers' marketing preferences in local coffee markets: Evidence from a choice experiment in Ethiopia. *Food Policy, 61*, 92–102.

Gibson, J. M., Rigby, D., Polya, D. A., & Russell, N. (2016). Discrete Choice Experiments in Developing Countries: Willingness to Pay Versus Willingness to Work. *Environmental & Resource Economics, 65*(4), 697–721.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science, 23*(3), 391–406.

Gilbride, T. J., & Allenby, G. M. (2006). Estimating heterogeneous EBA and economic screening rule choice models. *Marketing Science, 25*(5), 494–509.

González-Valdés, F., & Ortúzar, J. de D. (2018). The stochastic satisficing model: A bounded rationality discrete choice model. *Journal of Choice Modelling, 27*, 74–87.

González-Valdés, F., & Raveau, S. (2018). Identifying the presence of heterogeneous discrete choice heuristics at an individual level. *Journal of Choice Modelling, 28*, 28–40.

Goos, P., & Hamidouche, H. (2019). Choice models with mixtures: An application to a cocktail experiment. *Food Quality and Preference, 77*, 135–146.

Goos, P., Vermeulen, B., & Vandebroek, M. (2010). D-optimal conjoint choice designs with no-choice options for a nested logit model. *Journal of Statistical Planning and Inference, 140*(4), 851–861.

Gorton, M., & Barjolle, D. (2013). In *Theories of food choice* (pp. 15–26). Dordrecht: Springer.

Gracia, A., & Gómez, M. I. (2020). Food sustainability and waste reduction in Spain: Consumer preferences for local, suboptimal, and/or unwashed fresh food products. *Sustainability, 12*(10), 4148.
Green, P. E. (1974). On the design of choice experiments involving multi-factor alternatives. *Journal of Consumer Research, 1*(2), 61–68.
Green, P. E., & Srinivasan, V. (1990). Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice. *Journal of Marketing, 54*(4), 3–19.
Greene, W. H., & Hensher, D. A. (2003). A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transportation Research Part B Methodological, 37*(8), 681–698.
Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H., & Lancsar, E. (2021a). Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. *Journal of Choice Modelling, 100309*.
Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H., & Lancsar, E. (2021b). Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. *Journal of Choice Modelling, 100322*.
Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health Psychology, 28*(4), 404.
Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open, 5*(4), 1–9.
Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research, 67*, 1688–1699.
Hauser, J. R., Toubia, O., Evgeniou, T., Befurt, R., & Dzyabura, D. (2010). Disjunctions of conjunctions, cognitive simplicity, and consideration sets. *Journal of Marketing Research, 47*, 485–496.
Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics, 27*(1), 157–171.
Helter, T. M., & Boehler, C. E. H. (2016). Developing attributes for discrete choice experiments in health: A systematic literature review and case study of alcohol misuse interventions. *Journal of Substance Use, 21*(6), 662–668.
Hensher, D. A. (2006). How do respondents process stated choice experiments? – attribute consideration under varying information load. *Journal of Applied Economics, 21*(6), 861–878.
Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological, 44*(6), 735–752.
Hensher, D. A. (2014). Attribute processing as a behavioural strategy in choice making. In S. Hess, & A. J. Daly (Eds.), *Handbook of discrete choice modelling*. UK: Edward Elgar.
Hensher, D. A., Greene, W. H., & Chorus, C. G. (2013). Random regret minimization or random utility maximization: An exploratory analysis in the context of automobile fuel choice. *Journal of Advanced Transportation, 47*(7), 667–678.
Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation, 32*(3), 203–222.
Hess, S., & Hensher, D. A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological, 44*(6), 781–790.
Hess, S., Beck, M. J., & Chorus, C. G. (2014). Contrasts between utility maximisation and regret minimization in the presence of opt out alternatives. *Transportation Research Part A: Policy and Practice, 66*, 1–12.
Hoevenagel, R., & van der Linden, J. W. (1993). Effects of different descriptions of the ecological good on willingness to pay values. *Ecological Economics, 7*, 223–237.
Hoffman, S. D., & Duncan, G. J. (1988). Multinomial and conditional logit discrete-choice models in demography. *Demography, 25*(3), 415–427.
Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. *The Stata Journal, 7*(3), 388–401.
Holmes, T., Adamowicz, W., & Carlsson, F. (2017). Choice experiments. In: Champ, P.A., Boyle, K.J., & Brown, T.C. (eds). *A primer on nonmarket valuation*, Ch. 5. Springer, New York. pp. 133-186.
Horiuchi, Y., Markovich, Z. D., & Yamamoto, T. (2020). Does conjoint analysis mitigate social desirability bias? Available at SSRN: https://ssrn.com/abstract=3219323 or https://doi.org/10.2139/ssrn.3219323.
Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics, 69*, 1595–1603.
Jacquemet, N., Joule, R. V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management, 65*(1), 110–132.
Jaeger, S. R., Hort, J., Porcherot, C., Ares, G., Pecore, S., & MacFie, H. J. H. (2017). Future directions in sensory and consumer science: Four perspectives and audience voting. *Food Quality and Preference, 56*, 301–309.
Janssen, M., & Hamm, U. (2012). Product labelling in the market for organic food: Consumer preferences and willingness-to-pay for different organic certification logos. *Food Quality and Preference, 25*, 9–22.
Jarvis, W., Mueller, S., & Chiong, K. (2010). A latent analysis of images and words in wine choice. *Australasian Marketing Journal, 18*(3), 138–144.
Jedidi, K., & Kohli, R. (2005). Probabilistic subset conjunctive models for heterogeneous consumers. *Journal of Marketing Research, 17*, 483–494.
Jedidi, K., & Kohli, R. (2008). Inferring latent class lexicographic rules from choice data. *Journal of Mathematical Psychology, 52*, 241–249.
Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B. W., Kanninen, B., & Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in health, 16*(1), 3–13.
Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., & Tourangeau, R. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists, 4*(2), 319–405.
Johnston, R. J., Schultz, E. T., Segerson, K., Besedin, E. Y., & Ramachandran, M. (2012). Enhancing the content validity of stated preference valuation: The structure and function of ecological indicators. *Land Economics, 88*(1), 102–120.
Kaur, A., Scarborough, P., & Rayner, M. (2017). A systematic review, and meta-analyses, of the impact of health-related claims on dietary choices. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 1–17.
Kessels, R., Jones, B., & Goos, P. (2011). Bayesian optimal designs for discrete choice experiments with partial profiles. *Journal of Choice Modelling, 4*(3), 52–74.
Kessels, R., Jones, B., & Goos, P. (2015). An improved two-stage variance balance approach for constructing partial profile designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry, 31*(5), 626–648.
Kessels, R., Jones, B., & Goos, P. (2019). Using Firth's method for model estimation and market segmentation based on choice data. *Journal of Choice Modelling, 31*, 1–21.
Kessels, R., Goos, P., Jones, B., & Vandebroek, M. (2008). Recommendations on the use of Bayesian optimal designs for choice experiments. *Quality and Reliability Engineering International, 24*(6), 737–744.
Kessels, R., Palhazi Cuervo, D., & Sörensen, K. (2017). *Optimal design of discrete choice experiments with partial profiles involving a no-choice option, paper presented at the International Choice Modelling Conference*. South Africa: Cape Town.
King, M. F., & Bruner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology & Marketing, 17*(2), 79–103.
Klojgaard, M. E., Bech, M., & Sogaard, R. (2012). Designing a stated choice experiment: The value of a qualitative process. *Journal of Choice Modelling, 5*(2), 1–18.
Kohli, R., & Jedidi, K. (2005). Probabilistic subset conjunction. *Psychometrika, 70*(4), 737–757.
Kohli, R., & Jedidi, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science, 26*(3), 380–399.
Kragt, M. E. (2013). Stated and inferred attribute attendance models: A comparison with environmental choice experiments. *Journal of Agricultural Economics, 64*(3), 719–736.
Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865.
Kuhfeld, W. F., & Tobias, R. D. (2005). Large factorial designs for product engineering and marketing research applications. *Technometrics, 47*(2), 132–141.
Lagerkvist, C. J., Berthelsen, T., Sundström, K., & Johansson, H. (2014). Country of origin or EU/non-EU labelling of beef? Comparing structural reliability and validity of discrete choice experiments for measurement of consumer preferences for origin and extrinsic quality cues. *Food Quality and Preference, 34*, 50–61.
Lai, J., Wang, H. H., Ortega, D. L., & Widmar, N. J. O. (2018). Factoring Chinese consumers' risk perceptions into their willingness to pay for pork safety, environmental stewardship, and animal welfare. *Food Control, 85*, 423–431.
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157.
Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics, 26*(8), 661–677.
Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research, 61*(5), 534–547.
Le, T. T., Tran, T. T., Ho, H. T., Vu, A. T., & Lopata, A. L. (2018). Prevalence of food allergy in Vietnam: Comparison of web-based with traditional paper-based survey. *World Allergy Organization Journal, 11*(1), 1–10.
Lenk, P. J., DeSarbo, W. S., Green, P. E., & Young, M. R. (1996). Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs. *Marketing Science, 15*(2), 173–191.
Leong, W., & Hensher, D. A. (2012). Embedding decision heuristics in discrete choice models: A review. *Transport Reviews: A transnational Transdisciplinary Journal, 32*(3), 313–331.
Li, S., & Kallas, Z. (2021). Meta-analysis of consumers' willingness to pay for sustainable food products. *Appetite, 105239*.
Liebe, U., Glenk, K., Oehlmann, M., & Meyerhoff, J. (2015). Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys? *Journal of Choice Modelling, 14*, 17–31.
Liebe, U., Glenk, K., von Meyer-Höfer, M., & Spiller, A. (2019). A web survey application of real choice experiments. *Journal of Choice Modelling, 33*, Article 100150.
Liebe, U., Meyerhoff, J., & Hartje, V. (2012). Test–retest reliability of choice experiments in environmental valuation. *Environmental and Resource Economics, 53*(3), 389–407.
Liebe, U., Moumouni, I. M., Bigler, C., Ingabire, C., & Bieri, S. (2020). Using factorial survey experiments to measure attitudes, social norms, and fairness concerns in developing countries. *Sociological Methods & Research, 49*(1), 161–192.
Lindhjem, H., & Navrud, S. (2011). Using internet in stated preference surveys: A review and comparison of survey modes. *International Review of Environmental and Resource Economics, 5*, 309–351.
List, J. A., & Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics, 20*, 241–254.
Little, J., & Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: Further investigation using meta−analysis. *Economics Bulletin, 3*(6), 1–13.
Loomis, J. B. (2014). 2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics*, 34–46.
Louviere, J. J. (1984). Using discrete choice experiments and multinomial logit choice models to forecast trial in a competitive retail environment: A fast food restaurant illustration. *Journal of Retailing, 60*(4), 81–107.

Louviere, J., Hensher, D., & Swait, J. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press.
Loureiro, M. L., & Umberger, W. J. (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.
Lusk, J. L., & Schroeder, T. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics, 86*(2), 467–482.
Lusk, J. L., & Norwood, F. B. (2010). Direct versus indirect questioning: an application to the well-being of farm animals. *Social Indicators Research, 96*(3), 551–565.
Lusk, J. L., & Shogren, J. F. (2007). *Experimental auctions: Methods and applications in economic and marketing research*. Cambridge University Press.
Lusk, J. L., Jamal, M., Kurlander, L., Roucan, M., & Taulman, L. (2005). A meta-analysis of genetically modified food valuation studies. *Journal of Agricultural and Resource Economics*, 28–44.
Malone, T., & Lusk, J. L. (2018). A simple diagnostic measure of inattention bias in discrete choice models. *European Review of Agricultural Economics, 45*(3), 455–462.
Manski, C. F. (1977). The structure of random utility models. *Theory and Decisions, 8*, 229–254.
Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., Bredahl Jacobson, J., Liebe, U., Olsen, S. B., Sagebiel, J., & Thiene, M. (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis* (p. 129p). Springer Nature.
McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470.
Meissner, M., Pfeiffer, J., Peukert, C., Dietrich, H., & Pfeiffer, T. (2020). How virtual reality affects consumer choice. *Journal of Business Research, 117*, 219–231.
Meulders, M. (2013). An R Package for Probabilistic Latent Feature Analysis of Two-Way Two-Mode Frequencies. *Journal of Statistical Software, 54*(14), 1–29.
Meyerhoff, J., Oehlmann, M., & Weller, P. (2015). The influence of design dimensions on stated choices in an environmental context. *Environmental and Resource Economics, 61*(3), 385–407.
Michaels-Igbokwe, C., Lagarde, M., Cairns, J., & Terris-Prestholt, F. (2014). Using decision mapping to inform the development of a stated choice survey to elicit youth preferences for sexual and reproductive health and HIV services in rural Malawi. *Social Science & Medicine, 105*, 93–102.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine, 6*(7).
Mokas, I., Lizin, S., Brijs, T., Witters, N., & Malina, R. (2021). Can Immersive Virtual Reality Increase Respondents' Certainty in Discrete Choice Experiments? A Comparison with Traditional Presentation Formats. *Journal of Environmental Economics and Management (accepted for publication)*.
Møller, A. M., & Myles, P. S. (2016). What makes a good systematic review and meta-analysis? *British Journal of Anaesthesia, 117*(4), 428–430.
Mørkbak, M. R., & Olsen, S. B. (2015). A within-sample investigation of test–retest reliability in choice experiment surveys with real economic incentives. *Australian Journal of Agricultural and Resource Economics, 59*(3), 375–392.
Moser, R., Raffaelli, R., & Notaro, S. (2014). Testing hypothetical bias with a real choice experiment using respondents' own money. *European Review of Agricultural Economics, 41*(1), 25–46.
Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics, 30*(3), 313–325.
Nagashima, A. (1970). A comparison of Japanese and US attitudes toward foreign products. *The Journal of Marketing*, 68–74.
Nardi, V. A. M., Jardim, W. C., Ladeira, W., & Santini, F. (2019). Predicting food choice: a meta-analysis based on the theory of planned behaviour. *British Food Journal, 121*(10), 2250–2264.
Nguyen, T. T., Haider, W., Solgaard, H. S., Ravn-Jonsen, L., & Roth, E. (2015). Consumer willingness to pay for quality attributes of fresh seafood: A labeled latent class model. *Food Quality and Preference, 41*, 225–236.
Norwood, F. B., & Lusk, J. L. (2011). Social desirability bias in real, hypothetical, and inferred valuation experiments. *American Journal of Agricultural Economics, 93*(2), 528–534.
Oehlmann, M., Meyerhoff, J., Mariel, P., & Weller, P. (2017). Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management, 81*, 59–73.
Olsen, S. B. (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics, 44*(4), 591–610.
Onozaka, Y., & McFadden, D. (2011). Does local labeling complement or compete with other sustainable labels? A conjoint analysis of direct and joint values for fresh produce claims. *American Journal of Agricultural Economics, 93*(3), 689–702.
Paffarini, C., Torquati, B., Tempesta, T., Venanzi, S., & Vecchiato, D. (2021). Rural sustainability and food choice: the effect of territorial characteristics on the consumers' preferences for organic lentils. *Agricultural and Food Economics, 9*, 29.
Palma, D., Ortúzar, J. D. D., Rizzi, L. I., & Casaubon, G. (2018). Modelling consumers' heterogeneous preferences: a case study with Chilean wine consumers. *Australian Journal of Grape and Wine Research, 24*(1), 51–61.
Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*, 366–387.
Persson, J., & Wallin, A. (2015). The (misconceived) distinction between internal and external validity. In Against boredom: 17 essays on ignorance, values, creativity, metaphysics, decision-making, truth, preference, art, processes, Ramsey, ethics, rationality, validity, human ills, science, and eternal life to Nils-Eric Sahlin on the occasion of his 60th birthday. Fri tanke förlag. pp. 187-195.
Peschel, A. O., Grebitus, C., Colson, G., & Hu, W. (2016). Explaining the use of attribute cut-off values in decision making by means of involvement. *Journal of Behavioral and Experimental Economics, 65*, 58–66.
Peschel, A. O., Orquin, J. L., & Loose, S. M. (2019). Increasing consumers' attention capture and food choice through bottom-up effects. *Appetite, 132*, 1–7.
Prada, M., Godinho, C., Rodrigues, D. L., Lopes, C., & Garrido, M. V. (2019). The impact of a gluten-free claim on the perceived healthfulness, calories, level of processing and expected taste of food products. *Food Quality and Preference, 73*, 284–287.
Printezis, I., Grebitus, C., & Hirsch, S. (2019). The price is right!? A meta-regression analysis on willingness to pay for local food. *PLoS ONE, 14*(5).
Puckett, S. M., & Hensher, D. A. (2008). The role of attribute processing strategies in estimating the preferences of road freight stakeholders. *Transportation Research Part E: Logistics and Transportation Review, 44*, 379–395.
Rakotonarivo, O. S., Schaafsma, M., & Hockley, N. (2016). A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *Journal of Environmental Management, 183*, 98–109.
Ready, R. C., Champ, P. A., & Lawton, J. L. (2010). Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. *Land Economics, 86*(2), 363–381.
Reisch, L., Eberle, U., & Lorek, S. (2013). Sustainable food consumption: an overview of contemporary issues and policies. *Sustainability: Science, Practice and Policy, 9*(2), 7–25.
Rigby, D., Burton, M., & Pluske, J. (2016). Preference stability and choice consistency in discrete choice experiments. *Environmental and Resource Economics, 65*(2), 441–461.
Roberts, P., & Priest, H. (2006). Reliability and validity in research. *Nursing Standard, 20*(44), 41–46.
Rolfe, J., & Windle, J. (2015). Testing attribute selection and variation in a choice experiment to assess the tradeoffs associated with increased mining development. *Land Use Policy, 42*, 673–682.
Rousseau, S. (2015). The role of organic and fair trade labels when choosing chocolate. *Food Quality and Preference, 44*, 92–100.
Rousseau, S., & Vranken, L. (2013). Green market expansion by reducing information asymmetries: Evidence for labeled organic food products. *Food Policy, 40*, 31–43.
Ruseckaite, A., Goos, P., & Fok, D. (2017). Bayesian D-optimal choice designs for mixtures. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 66*(2), 363–386.
Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. Using discrete choice experiments to value health and health care. M. Ryan, K. Gerard and M. *Amaya-Amaya, Springer, Netherlands., 11*, 13–46.
Samuelson, W., & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1*, 7–59.
Sándor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research, 38*(4), 430–444.
Sandorf, E. D., & Campbell, D. (2019). Accommodating satisficing behavior in stated choice experiments. *European Review of Agricultural Economics, 46*(1), 133–162.
Sawada, M., Aizaki, H., & Sato, K. (2014). Japanese consumers' valuation of domestic beef after the Fukushima Daiichi Nuclear Power Plant accident. *Appetite, 80*, 225–235.
Scarpa, R., Ferrini, S., & Willis, K. (2005). Performance of Error Component Models for Status-Quo Effects in Choice Experiments. In R. Scarpa, & A. Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics* (Vol. 6, pp. 247–273). Netherlands: Springer.
Scarpa, R., Thiene, M., & Hensher, D. A. (2010). Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: does it matter? *Land Economics, 86*, 817–839.
Scarpa, R., Thiene, M., & Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics, 90*(4), 994–1010.
Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics, 95*, 165–180.
Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society: Series B (Methodological), 25*(2), 235–251.
Scheibehenne, B., Miesler, L., & Todd, P. M. (2007). Fast and frugal food choices: Uncovering individual decision heuristics. *Appetite, 49*, 578–589.
Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological bulletin, 134*(2), 207.
Shi, L., Xie, J., & Gao, Z. (2018). The impact of deal-proneness on WTP estimates in incentive-aligned value elicitation methods. *Agricultural Economics, 49*(3), 353–362.
Shimp, T. A., & Sharma, S. (1987). Consumer ethnocentrism: construction and validation of the CETSCALE. *Journal of Marketing Research, XXIV*, 280–289.
Siegrist, M., Ung, C. Y., Zank, M., Marinello, M., Kunz, A., Hartmann, C., & Menozzi, M. (2019). Consumers' food selection behaviors in three-dimensional (3D) virtual reality. *Food Research International, 117*, 50–59.
Simon, H. A. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics, 69*(1), 99–118.
Slade, P. (2018). If you build it, will they eat it? Consumer preferences for plant-based and cultured meat burgers. *Appetite, 125*, 428–437.
Sörqvist, P., Haga, A., Langeborg, L., Holmgren, M., Wallinder, M., Nöstl, A., Seager, P. B., & Marsh, J. E. (2015). The green halo: Mechanisms and limits of the eco-label effect. *Food Quality and Preference, 43*, 1–9.
Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: the food choice questionnaire. *Appetite, 25*(3), 267–284.

Street, D. J., & Burgess, L. (2007). *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. Hoboken, New Jersey: Wiley.

Sun, Y., & Latora, V. (2020). The evolution of knowledge within and across fields in modern physics. *Scientific Reports, 10*(1), 1–9.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological, 35*(10), 903–928.

Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological, 21*(2), 91–102.

Szolnoki, G., & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics and Policy, 2*(2), 57–66.

Teisl, M. F., Roe, B., & Hicks, R. L. (2002). Can eco-labels tune a market? Evidence from dolphin-safe labeling. *Journal of Environmental Economics and Management, 43*, 339–359.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology, 4*(1), 25–29.

Thurstone, L. L. (1931). The indifference function. *The Journal of Social Psychology, 2*(2), 139–167.

Traets, F., Sanchez, D. G., & Vandebroek, M. (2020). Generating optimal designs for discrete choice experiments in R: The idefix package. *Journal of Statistical Software, 96*(3), 1–41.

Train, K. E. (2001). A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit. *Working Paper, Department of Economics*. Berkeley: University of California.

Train, K. E. (2008). EM Algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling, 1*(1), 40–69.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.

Tully, S. M., & Winer, R. S. (2014). The role of the beneficiary in willingness to pay for socially responsible products: a meta-analysis. *Journal of Retailing, 90*(2), 255–274.

van Cranenburgh, S., & Chorus, G. C. (2018). Does the decision rule matter for large-scale transport models? *Transportation Research Part A: Policy and Practice, 114*, 338–353.

Van Loo, E. J., Caputo, V., Nayga, R. M., Jr, & Verbeke, W. (2014). Consumers' valuation of sustainability labels on meat. *Food Policy, 49*, 137–150.

Van Loo, E. J., Caputo, V., Nayga, R. M., Jr, Meullenet, J., & Ricke, S. (2011). Consumers' willingness to pay for organic chicken breast: Evidence from choice experiment. *Food Quality and Preference, 22*, 603–613.

Van Loo, E. J., Caputo, V., & Lusk, J. L. (2020). Consumer preferences for farm-raised meat, lab-grown meat, and plant-based meat alternatives: Does information or brand matter? *Food Policy, 95*(August).

van't Riet, J., Sijtsema, S. J., Dagevos, H., & De Bruijn, G.-J. (2011). The importance of habits in eating behaviour. An overview and recommendations for future research. *Appetite, 57*, 585–596.

Vanclay, J. K., Shortiss, J., Aulsebrook, S., Gillespie, A. M., Howell, B. C., Johanni, R., Maher, M. J., Mitchell, K. M., Stewart, M. D., & Yates, J. (2011). Customer response to carbon labelling of groceries. *Journal of Consumer Policy, 34*(1), 153–160.

Vecchio, R., Van Loo, E. J., & Annunziata, A. (2016). Consumers' willingness to pay for conventional, organic and functional yogurt: evidence from experimental auctions. *International Journal of Consumer Studies, 40*(3), 368–378.

Vermeulen, B., Goos, P., Scarpa, R., & Vandebroek, M. (2011). Bayesian conjoint choice designs for measuring willingness to pay. *Environmental and Resource Economics, 48*(1), 129–149.

Vermeulen, B., Goos, P., & Vandebroek, M. (2008). Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing, 25*(2), 94–103.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*(1), 213–239.

Vermunt, J. K., & Magidson, J. (2005). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.

Videbæk, P. N., & Grunert, K. G. (2020). Disgusting or delicious? Examining attitudinal ambivalence towards entomophagy among Danish consumers. *Food Quality and Preference, 83*.

Vlaeminck, P., Jiang, T., & Vranken, L. (2014). Food labeling and eco-friendly consumption: Experimental evidence from a Belgian supermarket. *Ecological Economics, 108*, 180–190.

Volinskiy, D., Adamowicz, W. L., Veeman, M., & Srivastava, L. (2009). Does choice context affect the results from incentive-compatible experiments? The case of non-GM and country-of-origin premia in canola oil. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 57*(2), 205–221.

Vondolia, G. K., & Navrud, S. (2019). Are non-monetary payment modes more uncertain for stated preference elicitation in developing countries? *Journal of Choice Modelling, 30*, 73–87.

Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical Social Sciences, 43*, 303–343.

Wang, J., Ge, J., & Ma, Y. (2018). Urban Chinese consumers' willingness to pay for pork with certified labels: A discrete choice experiment. *Sustainability, 10*(3), 603.

Wang, X., Yao, E., & Liu, S. (2018). Travel choice analysis under metro emergency context: Utility? Regret? Or Both? *Sustainability, 10*(11), 3852.

Wang, Z., An, S., Wang, J., & Ding, C. (2017). Evacuation travel behavior in regret minimization or utility maximization rules? Evidence from emergency context. *KSCE Journal of Civil Engineering, 21*(1), 440–446.

Weller, P., Oehlmann, M., Mariel, P., & Meyerhoff, J. (2014). Stated and inferred attribute non-attendance in a design of designs approach. *Journal of choice modelling, 11*, 43–56.

Wensing, J., Caputo, V., Carraresi, L., & Bröring, S. (2020). The effects of green nudges on consumer valuation of bio-based plastic packaging. *Ecological Economics, 178*.

Williamson, P. O., Lockshin, L., Francis, I. L., & Loose, S. M. (2016). Influencing consumer choice: Short and medium term effect of country of origin information on wine choice. *Food Quality and Preference, 51*, 89–99.

Williamson, P. O., Loose, S. M., Lockshin, L., & Francis, L. (2017). Predicting wine repurchase: a case of low test-retest reliability in China. *International Journal of Market Research, 59*(4), 471–494.

Windle, J., & Rolfe, J. (2011). Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis & Policy, 41*(1).

Wuepper, D., Clemm, A., & Wree, P. (2019). The preference for sustainable coffee and a new approach for dealing with hypothetical bias. *Journal of Economic Behavior & Organization, 158*, 475–486.

Yang, W., & Fang, L. (2020). Consumer Willingness to pay for food safety attributes in China: A meta-analysis. *Journal of International Food & Agribusiness Marketing*, 1–18.

Yang, W., & Renwick, A. (2019). Consumer willingness to pay price premiums for credence attributes of livestock products–A meta-analysis. *Journal of Agricultural Economics, 70*(3), 618–639.

Yeh, C. H., Hartmann, M., & Hirsch, S. (2018). Does information on equivalence of standards direct choice? Evidence for organic labels from different countries-of-origin. *Food Quality and Preference, 65*, 28–39.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science*, 579–616.

Zanoli, R., Scarpa, R., Napolitano, F., Piasentier, E., Naspetti, S., & Bruschi, V. (2013). Organic label as an identifier of environmentally related quality: A consumer choice experiment on beef in Italy. *Renewable Agriculture and Food Systems, 28*(1), 70–79.

Zawojska, E., & Czajkowski, M. (2017). Re-examining empirical evidence on stated preferences: importance of incentive compatibility. *Journal of Environmental Economics and Policy, 6*(4), 374–403.

Zeelenberg, M., & Pieters, R. (2007). A theory of regret regulation 1.0. *Journal of Consumer Psychology, 17*(1), 3–18.

Zeller, R. A., & Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge University Press.