European Review of Agricultural Economics Vol **50 (5)** (2023) pp. 1796–1823
doi:https://doi.org/10.1093/erae/jbad033
Advance Access Publication 3 October 2023

Comparing experimental auctions and real choice experiments in food choice: a homegrown and induced value analysis

Vincenzina Caputo[†,*], Angelos Lagoudakis[‡], Robert Shupp[†] and Claudia Bazzani[§]

[†]*Michigan State University, East Lansing, USA;* [‡]*Department of Economics, Iowa State University, Ames, Iowa, USA;* [§]*University of Verona, Verona, Italy*

Received June 2022; final version accepted September 2023

Abstract

This study compares a real choice experiment (RCE) with three commonly used experimental auction (EA) mechanisms (Becker–DeGroot–Marschak, random *n*th price auction (RNPA), second price auction (SPA)) to determine whether willingness to pay (WTP) estimates differ across these elicitation methods. We use quality labels on eggs as the empirical application and find that the SPA, RNPA and RCE yield similar WTP estimates, while the BDM mechanism generally produces higher WTP estimates. We also compare these EAs and the RCE in an induced value setting and find that the BDM auction produces greater deviations from the underlying value than the other EAs and RCE. We suggest that RCEs may be preferable to BDMs for collecting WTP estimates in logistically difficult experimental settings.

Keywords: willingness to pay, real choice experiment, BDM mechanism, random nth price auction, second price auction, induced value experiments

JEL Classification: C25, C57, D44, C91, C92, D11, D12

1. Introduction

Real choice experiments (RCEs) and experimental auctions (EAs) are two non-hypothetical non-market valuation methods commonly used to determine individuals' valuations for private goods. Most of the research using RCEs and EAs elicits homegrown values, which are the subjective evaluations of participants for a good, often expressed in terms of willingness to pay (WTP) values (Murphy, Stevens and Yadav, 2010).

*Corresponding author: E-mail: vcaputo@msu.edu

Unlike hypothetical stated preference methods like hypothetical choice experiments and contingent valuation, the EA and RCE methods are similar to active markets in that real money is exchanged to purchase actual goods (Lusk and Shogren 2007; Alfnes *et al.*, 2006). However, the way WTP values are elicited differs across these two methods. The EA method follows a direct approach, whereby consumers place their bids for a specific product without simultaneously evaluating other available products or product attributes. Once the experiment is completed, the bidder(s) buy the product, paying the price as defined by the auction mechanism. The RCE method, on the other hand, follows an indirect approach whereby participants are presented with repeated choice tasks (i.e. choice questions) and asked to make decisions (trade-offs) between product alternatives that are simultaneously offered at different posted prices. Once a respondent completes the experiment, one of these choice tasks is randomly selected as binding. The respondent is then required to buy the chosen alternative in the binding choice question and pay the posted price indicated in that alternative. The WTP values are then derived post-data collection through the estimation of discrete choice models (see Train, 2009).

Researchers are often confronted with a critical decision when they turn to experimental methods to elicit consumer demand for products and attributes: whether to use RCE or EA. From a theoretical perspective, both RCE and EA methods are considered incentive compatible; that is, participants are incentivized to truthfully reveal their preferences (Lusk, Feldkamp and Schroeder, 2004; Lusk and Schroeder, 2006). Hence, these methods should yield equivalent welfare estimates. However, research indicates that WTP estimates for private goods can differ across RCE and EAs, especially in food applications (e.g. Gracia, Loureiro and Nayga, 2011; Lusk and Schroeder, 2006; Shi, Xie and Gao, 2018; Cerroni *et al.*, 2019). Divergent results could be due to the diverse nature of the elicitation process underlying the RCE and EA methods. Indeed, previous studies indicate that direct and indirect preference elicitation methods yield different results when measuring attribute importance (e.g. choice-based conjoint, constant sum scales and best-worst scaling) (Louviere and Islam, 2008).

However, divergent bidding behaviours are also found across the three most popular EA mechanisms: the Becker–DeGroot–Marschak (Becker, DeGroot and Marschak, 1964), the second price auction (SPA) (Vickrey, 1961) and random nth price auction (RNPA) (Shogren *et al.*, 2001). Shogren *et al.* (2001) argued that RNPA might outperform SPA because the off-margin bidders (i.e. bidders whose values are not close to the second highest price) may not be de-motivated in bidding their true value for the good. Other authors suggest differences in bidding behaviour across SPA and BDM auctions may be due to psychological effects (including having a 'taste for winning') related to the group (SPA) versus individual (BDM) bidding settings (Lusk and Rousu, 2006; Canavari *et al.*, 2019). Generally, findings are mixed in terms of relative WTP values across EA mechanisms (List, 2003; Lusk, Feldkamp and Schroeder, 2004), although numerous studies indicate that people tend not to bid optimally in BDM auctions (see Canavari *et al.*, 2019 for a review). These last findings have prompted some authors to question the incentive compatibility of the BDM (Noussair, Robin and Ruffieux, 2004; Cason and Plott, 2014; Drichoutis and Nayga, 2022) and to identify potential sources of bias, including reference-dependent preferences and expectation-based reference points (see Vassilopoulos, Drichoutis and Nayga, 2018 for a detailed discussion). Nevertheless, the BDM auction remains the second most popular EA mechanism in the literature (Canavari *et al.*, 2019) likely due to its design features that make implementation easier compared to other EAs.

However, while the simplicity of implementing BDM auctions makes it attractive, the same design features can also have a significant impact on the results and conclusions drawn from BDM experiments. Therefore, further research is necessary to determine the effect of these logistical advantages on welfare estimates and, if needed, to suggest alternatives that present similar logistic benefits (Canavari *et al.*, 2019). This study adds to the existing non-market valuation literature by conducting a comparative analysis of the three most commonly used EA mechanisms (SPA, RNPA and BDM) with a RCE. Prior studies limited their comparisons of RCEs with only one type of EA mechanism; hence this study provides fresh insights into whether RCE and EAs differ consistently across alternative auction mechanisms. We employ a between-subject approach, conducting a homegrown RCE and three homegrown EAs—one for each EA mechanism. Participants were asked to express their preferences for a dozen large, grade A, brown eggs produced through either the conventional method, the cage-free system, or the USDA organic production requirements. Results indicate that the BDM mechanism produced higher estimates than the other mechanisms, while the SPA, RNPA and RCE did not generate statistically different WTP estimates from each other.

Our results from the homegrown experiments, while informative in terms of directional changes of welfare estimate across methods, remain inconclusive in determining which method most accurately reflects actual market behaviour. To further evaluate the demand-revealing properties of the EA mechanisms (SPA, RNPA and BDM) and RCE, this study also conducts induced value experiments (IVEs) of each as a robustness check. Unlike homegrown experiments, the IVEs involve both researchers and study participants being aware of the underlying value of the good under evaluation. Participants bid or choose based on the maximum payoff value (i.e. the IV), and researchers detect potential deviations from profit-maximizing behaviour (Smith, 1976). Previous studies have used IVEs to assess the validity of non-market valuation methods (Noussair, Robin and Ruffieux, 2004; Lusk and Rousu, 2006; Luchini and Watson, 2014; Cerroni *et al.*, 2019), but only Cerroni *et al.* (2019) have used food home-grown and IV experiments to explore differences in welfare estimates between RCE and one auction type, the SPA. By comparing IV-RCE with various IV-EA mechanisms, this study contributes to the limited but growing literature that uses IVEs as a demand-revealing validation process in non-market valuation studies.

The article is structured as follows: First, we provide some background on RCEs and EAs and discuss the existing literature. Next, we describe the data and sample. Subsequently, we report the experimental procedures and the econometric analyses implemented in the RCE and EAs. The last three sections present the results, the robustness checks and the conclusions.

2. Background

EAs and RCEs are widely used in many fields of applied economics, including agricultural and food economics (see Caputo and Scarpa, 2022 for a review on RCE in food choice; Canavari *et al.*, 2019 for a review of EAs). The popularity of these methods is partially because their ability to mitigate hypothetical bias[1] in welfare estimates while also showing a high level of external validity (Lusk and Schroeder, 2004; Chang, Lusk and Norwood, 2009; Brooks and Lusk, 2010).

2.1. Incentive compatibility and experimental auctions

EAs constitute a direct preference elicitation method given that the bids subjects place for a specific product represent a direct measure of participants' WTP for that product. When implementing EAs, researchers need to make several implementation decisions, including which EA mechanism to implement. The selection of EA mechanisms depends on a range of factors, including logistical considerations and convenience.

In the agricultural economics literature, the most used EA mechanisms are SPA, BDM and RNPA (Canavari *et al.*, 2019). These auctions vary in terms of the number of individuals participating simultaneously in the experiment and how the market clearing price is determined. With regard to participant numbers, the SPA and RNPA are conducted with groups of more than two individuals, while the BDM can be conducted with one individual at a time. The market clearing price in the SPA and RNPA is determined by the highest bids submitted by the participants, while in the BDM auction, the market price is drawn from a random distribution. These different setups have been shown to induce different behaviors[2] and welfare estimates across mechanisms (Rutström, 1998; Lucking-Reiley, 1999; List, 2003; Lusk, Feldkamp and Schroeder, 2004).

This has raised questions about the incentive compatibility of the various EA mechanisms. For SPA and RNPA auctions, the foundation for their incentive compatibility lies in the fact that participants have a weakly dominant strategy, meaning that their strategy yields a payoff at least as high as any other strategy, regardless of what the other participants in the auction do. According to expected utility theory (EUT), three key assumptions must hold: first, each individual's private value must be drawn from a distribution that is common

1 See recent studies regarding hypothetical bias in non-market valuation methods by Penn and Hu (2018) and Haghani *et al.* (2021).

2 A number of studies explored whether and how behavioural factors characterizing the participants influence behaviours. These factors include cognitive ability, emotions, mood, altruism and spite. Some of these factors could be particularly influential in SPA and RNPA as participants compete against each other (see Canavari *et al.*, 2019 for a comprehensive review on these factors).

knowledge and independent of the other participants' values; second, only one divisible product is available for sale (e.g. eggs in our case); and lastly, participants are assumed to have a differentiable utility function, and their valuations of risky outcomes are defined by the EUT (Lusk and Shogren 2007). As shown in Lusk and Shogren (2007), an expected utility model can be used to describe the bidding behaviour of a participant, n, who submits a bid, b_n, against N rival bidders to maximize their expected utility:

$$E[U_n] = \int_{\underline{p_n}}^{b_n} U_n(v_n - p)\, dG_n(p) + \int_{b_n}^{\overline{p_n}} U_n(0) \tag{1}$$

$$= \int_{\underline{p_n}}^{b_n} U_n(v_n - p)\, g_n(p)\, dp + \int_{b_n}^{\overline{p_n}} U_n(0)$$

where v_n represents the individual's value for the product, p is the market price, U is a utility function increasing in income, and $G_n(p)$ is the cumulative distribution function (CDF) which characterizes the bidders' expectation about the price. The CDF has support $\left[\underline{p_n}, \overline{p_n}\right]$, and $g_n(p)$ is the associated probability density function.

As argued by Lusk and Shogren (2007), the highest bidder's derived utility is the difference between his/her value and the market price (second highest bid or random nth highest bid). Every other bidder's (non-winning) value from bidding gets normalized to zero. As can be seen in Eq. (1), the first integral is calculated over all prices smaller than the individual's bid (winning cases), and the second integral is calculated over all prices higher than the individual's bid (non-winning cases). After normalizing $U(0) = 0$, the optimal bid is found by taking the derivative of Eq. (1) with respect to the participant's bid, b_i, and setting the derivative equal to zero as follows: $\frac{\partial E[U_n]}{\partial b_n} = U_n(v_n - b_n)\, g_n(b_n) = 0$, which is solved when $b_n = v_n$. Hence, we can conclude that the individual's expected utility is maximized when he/she submits a bid equal to his/her value for the good. According to Lusk and Shogren (2007), this finding does not depend on the bidder's risk preferences, number of bidders, initial endowment or other bidders' bidding strategies.

The BDM mechanism is often considered incentive compatible, but this has been questioned by several authors due to a number of issues.[3] Banerji and Gupta (2014) note significant differences in BDM valuations when the range of the randomly drawn prices is changed (in accordance with expectation-based reference points). Cason and Plott (2014) argue that the BDM mechanism is problematic due to the difficulty of understanding the mechanism. Vassilopoulos, Drichoutis and Nayga (2018) indicate that bids generated from the BDM mechanism are dependent on the anchoring of bids to the chosen

3 Due to these issues, BDM is typically referred to as a mechanism and not as an auction. For exposition reasons, we refer to it as an auction.

price support. Most notably, Karni and Safra (1987) proved that if the independence axiom of EUT is relaxed, the optimal bid in a BDM mechanism will not necessarily be equal to the certain equivalent of the lottery.

To illustrate, the independence axiom states that if an individual prefers lottery **A** to lottery **B**, that individual will also prefer the combination $\alpha A+(1-\alpha)C$ to $\alpha B+(1-\alpha)C$ for all α between zero and one and for any **C.** The axiom dictates this linear-in-probabilities property (Machina, 1982). When individuals are bidding in the BDM mechanism, they are evaluating a compound lottery (made up of the lottery itself and the randomly generated price). Assuming people are not weighting probabilities linearly, they violate the independence axiom and hence, EUT. In that case, the value of the compound lottery differs from the underlying lottery of interest (BDM mechanism). Hence, the incentive compatibility of the mechanism is questioned.

2.2. Experimental auctions versus real choice experiments

RCEs are similar to BDM auctions in that experiments can be conducted with one individual at a time. According to Lusk and Schroeder (2004), RCE are more flexible than EAs because (i) the evaluation of product alternatives or food attributes occurs simultaneously[4] and (ii) the choice tasks (i.e. choice questions) are designed in a way that more closely mirrors actual shopping situations (e.g. making a choice among multiple products offered at different prices). However, unlike EAs, the WTP values are elicited through the estimation of econometric models (an indirect approach).

A number of food choice studies have compared RCEs with some EAs and found that they may yield different outcomes in terms of welfare estimates such as WTP (see Table A1 in Appendix A, supplementary data at ERAE online). Lusk and Schroeder (2006) investigated whether WTP for steak attributes differ between a SPA and an RCE. Using a between-sample approach, consumers were asked to participate either in an auction market (SPA auction) or in a RCE. Findings from this study can be summarized as follows: (a) auction bids were significantly lower as compared to WTPs from the RCE, (b) own-price elasticities of demand for higher quality products were notably higher when derived from the auction data than when derived from the RCE data and (c) the consumers' preference orderings were similar across the two elicitation methods. In the same vein, Gracia, Loureiro and Nayga (2011), using a between-sample approach, compared RCE and RNPA using cured-ham products, differentiated by four different levels of an animal welfare label. The authors concluded that WTP estimates under RNPA and RCEs have the same sign (positive) but different magnitudes, especially across individual-specific socio-demographic characteristics.

Based on the results of these earlier studies, subsequent research has attempted to explore factors driving variations across RCE and EAs. Grebitus,

4 In the RCEs, products are described by multiple attributes that participants can simultaneously evaluate. In contrast, in the EAs, products are typically evaluated one-by-one, and the number of attributes included in the design is minimized to facilitate operations.

Lusk and Nayga (2013) conducted real and hypothetical SPAs and RCEs for apples and wine. Their results suggest that differences in choice outcomes between RCE and SPA are driven by subjects' personality traits. Shi, Xie and Gao (2018) conducted three experiments: an RCE, a real double-bounded dichotomous contingent valuation (RCVM), and a BDM auction on three different types of orange juice. The authors found that lower levels of deal-proneness led to smaller differences in WTP estimates across the BDM auction compared to RCE and RCVM. This suggests that the 'gambling behaviour' of deal-prone individuals may be prompted by the BDM auction mechanism, resulting in understated bids.

More recently, Cerroni *et al*. (2019) combined induced value and homegrown procedures for RCE and SPA to explore whether the differences between these two mechanisms exist because of different levels of demand revelation or because of differences in value-formation in homegrown preferences. They found that homegrown preference patterns are different between the two elicitation methods. In addition, their investigation of IV preferences revealed that RCE was the most demand-revealing elicitation method.

3. Procedures and implementation

Upon arriving at the lab for a session, participants were randomly assigned to one of the four experiments: SPA, RNPA, BDM or RCE. Following the standard practice, each experiment consisted of multiple sessions. The BDM and RCE experiments were conducted in one-on-one interviews, and the SPA and RNPA experiments were conducted in groups of five. The SPA and RNPA experiments were randomized at the session level, as recommended by Lonati *et al*. (2018). In addition, as suggested by Canavari *et al*. (2019), they were executed with a small and constant number of participants (five) per session to keep everyone equally engaged and to mitigate potential confounding social dynamic effects on bidding behaviour. This set up was also motivated by recent theoretical (Banerji and Gupta, 2014) and empirical studies (Rosato and Tymula, 2019), which indicate that when the number of bidders per session increases, equilibrium bids tend to decrease.

In all homegrown experiments, each session was performed in seven phases: consent form and participation fee, a short survey, product display and information, instruction, practice rounds, the experiment, and a second short survey. Participants were first given a cash participation fee of $13 and asked to read and sign a consent form. Next, they completed a short questionnaire of general demographic and consumption questions. Participants were then presented with a display table featuring three types of eggs:[5] conventional, USDA organic and cage free. A captioned picture describing each type of egg was read aloud and shown to participants via a PowerPoint presentation (see Figure B1

5 Eggs were selected in this study because they are widely available in various food outlets including grocery stores, convenience stores and farmers' markets all over the United States. All eggs were of similar size (1-dozen), grade (A), colour (brown) and packaging and non-branded products were used to avoid potential branding effects.

in Appendix B). After viewing the eggs, participants proceeded through the instructions and practice rounds and then participated in the actual experiment. After the completion of the experiment, participants completed a second short survey.

The following subsection describes the procedures followed to implement the EAs and RCE in more detail.

3.1. Experimental auctions

Participants who were randomly selected to participate in one of the three EAs (SPA, BDM and RNPA) were first presented with a hypothetical auction for four different candy bars to familiarize them with the procedure of their auction type. After completing the practice auction, they participated in an auction for each type of egg (conventional, cage-free and USDA organic). To avoid ordering effects, which have been previously noted by Demont *et al*. (2012), the order in which the three types of eggs were displayed and auctioned was randomized in all EAs. The experimental procedures for the SPA, RNPA and BDM auction are outlined below.

- *Step 1*. A total of three bidding rounds were conducted, one for each type of egg product. At the beginning of each round, the participant(s) received a bid sheet and were asked to bid (simultaneously, if in a group) for the product being auctioned in that round. The bidding sheets were collected before moving on to the next egg product.
- *Step 2*. The experimenter rolled a die to determine which egg auction was binding. Importantly, all the auctions had an equally likely chance of being binding.
- *Step 3*. For the SPA and the RNPA, the bids in the chosen auction were confidentially ranked from highest to lowest. In the SPA, the person with the highest bid purchased the eggs but paid the second highest bid. In the RNPA, a random number (N) was drawn by rolling a die to determine how many participants would purchase/win the eggs. The random number was somewhere between 2 and 5 (number of participants). The N-1 highest bidders in the binding egg auction purchased the eggs and paid the nth highest bid. In the BDM auction, the experimenter rolled a 10-sided die two times (one for the second decimal and one for the first decimal) and a 7-sided die one time to determine a randomly drawn price between $0.00 and $6.00.[6] If the bid for the binding eggs was greater than or equal to the randomly drawn price, the participant purchased the eggs and paid the randomly drawn price. If the bid for the binding eggs was less than the randomly drawn price, the participant paid nothing and received nothing.
- *Step 4*. In the SPA and the RNPA, for the chosen egg auction, the experimenter wrote the winning bidder(s) number and the price paid (second highest bid/nth highest bid) on the board for everyone to see. In the BDM

6 If the generated price was above $6.00 after rolling the 7-sided die, we rolled it again until the price was within the predetermined range [$0.00, $6.00].

auction, for the chosen egg auction, the experimenter wrote the randomly drawn price (between $0.00 and $6.00) on the board. At the end of the experiment, the winning bidder(s) paid the second highest bid/random nth highest bid/randomly drawn price (for the SPA, the RNPA and the BDM auction respectively) and obtained the eggs. All other participants paid nothing and received nothing.

3.2. Real choice experiment

In the RCE, participants were presented with a series of repeated choice tasks (or choice questions), each including three alternatives: two types of eggs and a no-buy ('none of these') option. For each choice task, participants had to select one of the two egg products or choose not to buy any. The eggs used in the experiment were a dozen large, grade A, brown eggs and were described by three attributes and their respective levels: price ($1.59, $2.59, $3.59 and $4.59), USDA-organic label (present/absent) and cage-free label (present/absent). The price levels for the eggs were selected based on the prices at local grocery stores and retail prices reported by the U.S. Department of Agriculture—Agricultural Marketing Service and the USDA's National Retail Reports at the time of the experiment.

To determine the number of choice tasks presented to participants, a D-Optimal design proposed by Street and Burgess (2007) was used,[7] which resulted in eight choice tasks with a D-Optimality of 96.6 percent. These eight choice tasks were randomly divided into two blocks of four tasks each, and each participant was assigned only four tasks. The order in which the choice tasks were presented to respondents in each block was randomized at the individual level to prevent any ordering effects. An example of a choice task is provided in Figure B2 in the Appendix B (supplementary data at ERAE online).

To implement the RCE, we followed established protocols from previous studies (Lusk and Schroeder, 2006; Gracia, Loureiro and Nayga, 2011; Bazzani *et al.*, 2017) and conducted individual sessions with each participant. To familiarize participants with the RCE procedures, we first presented them with four hypothetical choice tasks involving candy bar selection. Afterward, they participated in the RCE for egg selection, which consisted of three steps:

- *Step 1:* Each participant was given a choice sheet and presented with four choice tasks, one at a time. Each task required the participant to select their preferred egg product at the listed price or to choose the no-purchase option. Participants recorded their choices on the provided choice sheet.
- *Step 2:* After the participant had completed the four choice tasks, the experimenter collected the choice sheet.
- *Step 3:* The experimenter rolled a four-sided die to determine which of the completed choice tasks would be binding. For example, if a one was rolled,

7 As in Gracia, Loureiro and Nayga (2011) and consistent with our experimental setup for the home-grown EAs, our RCE experimental design allows only for the estimation of main effects. No interaction terms were included in the experimental design.

the first-choice question was binding, and if the participant had chosen one of the two types of eggs, he/she was given the product selected and was asked to pay the price listed in the choice. If the participant had chosen the 'no-purchase' option, then received no eggs and paid nothing.

3.3. Sample characteristics

We recruited 270 participants[8] from a random sample of 2000 consumers pulled from the general population of East Lansing and Lansing, Michigan, United States. Participants were recruited via email using lists managed by Michigan State University. The emails stated that participants would be paid $13 to participate in a 45-min study on egg consumption; no other information about the experiment was provided to the participants. Only participants older than 18 years, responsible for grocery shopping, not lactose intolerant, and not following a vegan diet were selected for the study.

We implemented a between-sample approach,[9] where selected participants were randomly allocated to one of the four experiments. A total of 69, 70, 69 and 62 participants participated in the SPA, BDM, RNPA and RCE experiments, respectively. Each experiment consisted of multiple sessions scheduled Monday through Sunday during morning, afternoon, and evening hours to account for timing effects (Lusk and Shogren 2007). Table A2 (online supplementary material) provides summary statistics and variable definitions of the basic demographic, consumption habits, and experimental-related questions for all the treatments. We tested the null hypothesis of equality of means (one-way analysis of variance) across treatments at the 5 per cent significance level and found no significant differences in demographic characteristics, except for the years of education. Descriptive statistics for consumption habits and experimental-related variables were also balanced.

4. Econometric analysis

4.1. Experimental auction models and specification

In the EAs, participants were required to submit simultaneous bids for each one of the three types of eggs: conventional, cage free and USDA organic eggs. The actual bids submitted by participants in each auction represent their total WTP for the respective egg types. These actual bids were then used to explore whether there exists a statistically significant difference in actual bids (or total WTP) elicited from the three different EA mechanisms. We employed two empirical strategies for this purpose. The first strategy involved using F-tests and post-hoc pairwise comparisons. The following null hypothesis was tested:

8 A total of 279 consumers participated in our experiments. At the beginning of our data input/analysis, we dropped nine observations due to incomplete/inattentive responses: seven observations from the RCEs, one observation from the SPA and one observation from the RNPA.

9 We employed a between-subject approach to avoid potential biases introduced by participation in multiple experiments (Lusk and Schroeder, 2006), fatigue effects (Charness, Gneezy and Kuhn, 2012) and the potential reduction in demand when consumers purchase multiple products (Lusk and Schroeder, 2006).

$H_0 = tWTP^{SPA} = tWTP^{BDM} = tWTP^{\mathrm{RNPA}}$. Failing to reject this hypothesis, we would conclude that there is statistical equality among the total WTPs elicited from the three egg types.

The second strategy relies on the estimation of a random effect[10] Tobit model. The Tobit model is the most widely used econometric method to analyse EA data, especially when there exists a significant number of zero bids (see Canavari *et al.*, 2019). In our case, a total of 20 participants placed zero bids: 8, 5 and 7 for the SPA, BDM and RNPA, respectively (see Table A3 in Appendix A, supplementary data at ERAE online). In addition, this model is preferred because in EAs actual bids (or total WTP) are always greater than or equal to zero, as also pointed out by Gracia, Loureiro and Nayga (2011) and Lusk and Schroeder (2006). We estimated four pooled Tobit models:

$$T\_WTP_{nt} = \alpha' x_{nt} + u_n + \upsilon_{nt} \tag{2}$$

where the independent variable, T_WTP_{nt}, is the individual n's actual bids (or total WTPs) for a different type of egg (i.e. USDA Organic and Cage-Free) at each time period t.x_{nt} is a vector of independent variables; u_n is an individual specific disturbance for participant n; and υ_{nt} is the overall error term.

The four models differed in the independent variables (x_{nt}) included in the estimation process. Pooled Tobit 1 includes the EAs mechanisms and the products (cage free and USDA organic) as independent dummy variables. Pooled Tobit 2 adds to Pooled Tobit 1 by also including interaction terms between the auction mechanisms and the type of eggs as independent variables. Pooled Tobit 3 adds to Pooled Tobit 2 by incorporating demographics as independent variables. Pooled Tobit 4 adds to Pooled Tobit 3 by including consumption and experiment-related variables as additional independent variables.

4.2. Real choice experiment models and specification

We estimated the RCE data using a mixed logit model (MXL) for panel data which allows for random state variation, unrestricted substitution patterns, and correlation in unobserved factors over time (Train, 2009).[11] Consider a sequence of observed choices **i** by individual n, one for each time period t (i.e. choice task), the unconditional probability that individual n makes this sequence of choices is represented by:

$$P_{n\mathbf{i}} = \int \prod_{t=1}^{T} \left[\frac{exp\left(V_{njt}\right)}{\sum_j exp\left(V_{njt}\right)} \right] f(\beta)\, d\beta. \tag{3}$$

10 We incorporated random effects into the Tobit models in order to account for the panel nature of the data (i.e. each participant submitted one bid for each type of eggs: conventional, USDA organic, cage free).

11 Given our limited sample size, we chose to use the MXL model in preference space instead of the MXL model in WTP space. We made this decision to prioritize a model that was more parsimonious and reliable. In fact, we found that the MXL model in WTP space had convergence issues at Halton draws below 2,000. In addition, model fit criteria such BIC and AIC suggested that the MXL model in WTP space underperformed compared to the MXL model in preference space.

The deterministic component of the utility, V_{njt}, was specified as:

$$V_{njt} = ASC_{none} + \alpha_{price} * PRICE_{njt} + \beta_{n,FREE} * CFREE_{njt} + \beta_{n,USDA} * USDA_{njt} \tag{4}$$

where ASC_{none} is an alternative specific constant representing the no-purchase option; $PRICE_{njt}$ is the price (continuous) variable generated by the price levels in our experimental design; $USDA_{njt}$ and $CFREE_{njt}$ are dummy variables for the USDA organic and cage free labels. They take a value of 1 when the label is present in the product, and 0 otherwise; α_{price} is the price coefficient and β_{FREE} and β_{USDA} are the coefficients for the cage-free and USDA organic labels. The parameters of the model are estimated by simulated maximum likelihood estimation techniques, using 500 Halton draws. The random coefficients for the USDA organic and the cage free labels are assumed to follow a normal distribution, while the price coefficient is assumed to be fixed.

4.3. Comparison of experimental auctions and real choice experiment: modelling approach and specification

Following the approach by Lusk and Schroeder (2006), we converted the RCE data into EA (continuous) data and focused on comparisons between EAs and RCE based on *m*WTP values.[12] In the SPA, BDM and RNPA auctions, the *m*WTP for USDA organic/cage free eggs was calculated by subtracting the bids for USDA organic/cage free eggs from the bids for conventional eggs. In the RCE, on the other hand, we derived individual-specific *m*WTPs for each egg type by applying the conditional inference procedure as described by Train (2009).

We then compared *m*WTP for USDA organic and cage free eggs obtained from the four homegrown experiments (SPA, RNPA, BDM and RCE) following two approaches. The first approach consisted of computing descriptive statistics (means) and testing the null hypothesis of equality of mean *m*WTP across experiments (SPA, RNPA, BDM and RCE). We did so by using parametric and non-parametric tests (T-test, Wilcoxon test, Poe, Giraud and Loomis and 2005[13] test).

The second approach relies on the use of a segmented sample approach and estimation of three Tobit random-effects models, one for each experiment (SPA, RNPA and BDM). For each EA, the Tobit model was specified as in equation (2): $T\_WTP_{nt} = \alpha' x_{nt} + u_i + \upsilon_{nt}$. The dependent variable (T_WTP_{nt}) is the total WTP collected from each auction mechanism for each egge type; x_{nt} is a vector of independent dummy variables that identify USDA organic and Cage-Free eggs, while conventional eggs is the baseline. Hence, the vector of coefficients, α', represents the $mWTP$ values for USDA organic and Cage-Free eggs versus conventional eggs.

12 It is important to note that EA data is continuous, while data from the RCE is discrete. To compare EAs with RCEs, the EA data can be converted into RCE data, and vice versa, the RCE data can be transformed into continuous data (see Lusk and Schroeder 2007 for more details).

13 The Poe et al. test was performed using 1,000 bootstrapped mWTP values generated from the mean *m*WTP values for each product type and their respective standard errors.

The estimated coefficients and respective variance-covariance matrix from each of the three Tobit models were used to generate a distribution of 1,000 *mWTP* for each product type (USDA Organic and Cage-Free eggs) using the parametric bootstrapping method introduced by Krinsky and Robb (1986). The same procedure was applied for the RCE data using the estimated coefficients and variance-covariance matrix from the MXL model. This process yield 1,000 bootstrapped values of *mWTP* for each product type and experiment. The bootstrapped values were used to perform the nonparametric test proposed by Poe, Giraud and Loomis (2005), testing the null hypothesis of equality of *mWTP* for USDA organic and cage-free eggs between homegrown experiments.

5. Results

5.1. Experimental auctions estimates

Table A4 in Appendix A (see supplementary data at ERAE online) reports the actual auction bids (or total WTP) for a dozen of each egg type, segregated by auction treatment and egg type.[14] The last columns of the table report the results from parametric and non-parametric hypothesis tests. These indicate that actual bids are statistically equivalent across SPA and RNPA auctions. However, in the BDM auction, bids are significantly higher for USDA organic ($2.042) and cage-free ($1.726) eggs, compared to the SPA ($1.592 and $1.148 for USDA organic and cage free eggs, respectively) and RNPA ($1.556 and $1.262 for USDA organic and cage free eggs, respectively) auctions. No significant differences in actual bids are found for conventional eggs across EAs.

To obtain a more complete picture of the auction bids from the three EA mechanisms, it may be useful to consider the entire distribution of WTPs, as also suggested by Lusk and Schroeder (2006). Figures B3 through B5 in Appendix B (supplementary data at ERAE online) show the inverse CDFs of actual bids for the conventional, USDA organic and cage free eggs, respectively. The CDFs can be interpreted as demand curves, assuming that each

14 We based our comparison of the actual bids on three assumptions that were derived from the study by Alfnes (2009). Our first assumption states that the participants in the study knew the value of eggs (market prices) and that the expected value of the market options was the same for all egg types auctioned. Hence, the effect of market price cancelled out when comparing the actual WTP bids. We can assert that this assumption holds true since the auctioned products had very similar characteristics and were close substitutes that could potentially replace the same egg product from the market. Our second assumption assumed that the studied participants had an underlying value for eggs. This is a reasonable assumption since all the participants were egg consumers, and eggs are a product with a relatively stable demand curve. Our third assumption is more general and postulates that consumers' valuations for different egg alternatives are not context-dependent and are based on a utility maximization process under a budget constraint. In our study, participants could only obtain one dozen of eggs, and they were all given the same endowment of $13. Additionally, during the implementation of our experiments, we attempted to maintain a high level of experimental control by preventing participants from communicating with each other and isolating them from outside clues. We also acknowledge there are many factors that could potentially affect participants' bids, and further research is necessary to evaluate the robustness and external validity of our findings.

participant only purchases (consumes) one unit and no other product alternative exists in the market for each figure. The results show that the distributions of WTP obtained from the BDM auction tend to lie above those of the SPA and the RNPA. Consistent with our earlier findings, the CDFs indicate that the BDM auction derives, on average, a higher WTP than the SPA and the RNPA for all types of eggs. In addition, with the exception of cage-free eggs (where RNPA results in slightly higher WTPs than the SPA), we can conclude that the SPA and the RNPA derive similar WTPs.

To further determine the effect of auction institution on WTP estimates, we estimated four pooled random-effects Tobit models for panel data.[15] As mentioned earlier, Pooled Model 1 allows us to explore the effect of EA on bids for USDA and cage free eggs, ignoring the other experimental factors. The dependent variable is the actual bids (total WTP) from the different auctions, and the independent variables are USDA organic and cage free eggs, while conventional eggs are the baseline. Pooled Model 2 includes the set of interaction terms between the egg types and the EA types as independent variables in addition to the variables in Pooled Model 1. Pooled Model 3 further adds socio-demographic variables as independent variables, and Pooled Model 4 incorporates consumption and experimental-related variables as independent variables, including purchase frequencies, expected prices, familiarity with the cage-free and USDA organic labels, and the time the experiment took place (morning versus afternoon). Table 1 presents the average (conditional marginal) effects on the observed censored variables from the four pooled Tobit models. We treated SPA and conventional eggs as baselines for identification purposes. The demographics and consumption/experiment-related variables included in models 3 and 4 are described in Table A2, supplementary data at ERAE online.

Results reveal that the BDM yields higher WTP estimates for both the cage free and USDA organic egg versus conventional eggs, as evidenced by the positive and statistically significant coefficients of the BDM variable across the four models ($0.428, $0.432, $0.440 and $0.380 in the pooled models 1–4, respectively). Interestingly, the coefficient of the RNPA is not statistically significant in any of the four models, indicating that the SPA and RNPA produce equivalent WTPs. Even after adding interaction terms between EA type and egg type (Pooled Model 2), socio-demographic variables (Pooled Model 3), and other consumption and experimental related variables (Pooled Model 4), the main result remains unchanged: WTP for both USDA organic eggs is higher than conventional eggs when the auction mechanism implemented

15 In panel data analysis, Tobit models may be influenced by deviations from normality and homoscedasticity, potentially affecting their accuracy. To address this issue, we estimated heteroscedastic Tobit models and compared them with the homoscedastic Tobit models reported in Table 1. For the heteroscedastic Tobit models, we followed the approach proposed by Shehata (2011). To determine which model provided the best fit, we conducted a likelihood ratio (LR) test. Our findings indicated no significant differences between the heteroscedastic and homoscedastic Tobit models.

Table 1. Estimates from the Tobit models, average marginal effects

Dependent variable	Total WTP			
Independent variables	Pooled Model 1	Pooled Model 2	Pooled Model 3	Pooled Model 4
Main effects				
BDM vs. SPA	0.428***	0.432***	0.440***	0.380**
	(0.159)	(0.159)	(0.158)	(0.158)
RNPA vs. SPA	0.039	0.038	0.018	0.024
	(0.152)	(0.152)	(0.152)	(0.151)
USDA organic vs. conventional	0.739***	0.741***	0.741***	0.740***
	(0.062)	(0.062)	(0.062)	(0.062)
Cage free vs. conventional	0.400***	0.403***	0.403***	0.403***
	(0.059)	(0.059)	(0.059)	(0.059)
Female			0.043	0.056
			(0.137)	(0.136)
Age			−0.005	−0.004
			(0.005)	(0.005)
Education			0.048*	0.044
			(0.027)	(0.027)
Income			0.039	0.045*
			(0.026)	(0.026)
Household size			0.010	−0.008
			(0.058)	(0.058)
Time experiment				0.200
				(0.137)
Expected price				0.081
				(0.049)
Familiarity cage free				0.100
				(0.132)
Familiarity USDA				−0.005
				(0.227)
Purchase frequency				0.148
				(0.145)
Interaction effects				
USDA Organic × BDM		0.469**	0.478**	0.411**
		(0.197)	(0.197)	(0.196)
Cage Free × BDM		0.590***	0.598***	0.536***
		(0.184)	(0.183)	(0.183)
USDA Organic × RNPA		−0.015	−0.037	−0.030
		(0.191)	(0.191)	(0.190)
Cage Free × RNPA		0.112	0.091	0.099
		(0.175)	(0.175)	(0.174)
Log likelihood	−823.05	−820.63	−817.92	−814.62
N. of groups	208	208	208	208

Note: Standard errors are in the parentheses. Single asterisk (*), double asterisk (**) and triple asterisk (***) denote the significance level at 10 per cent, 5 per cent and 1 per cent, respectively.

is the BDM.[16] These results align with previous studies, such as Shogren *et al*. (1994), who found no statistically significant differences in mean WTP estimates between SPA and RNPA. Our results also support the findings of Noussair, Robin and Ruffieux (2004), who found that SPA auctions produce more reliable WTP estimates compared to the BDM auction, and Lusk and Rousu (2006), who found that the SPA and RNPA auctions produce more reliable WTP estimates than the BDM.

Based on the empirical results presented in Table 1, it is also evident that consumers are WTP a higher price premium for USDA organic eggs compared to cage free eggs. National statistics confirm this empirical result, as the advertised average price for USDA organic eggs in major retail supermarket outlets during 2018 was $0.28 higher than that for cage free eggs (USDA National Retail Report—Shell Egg and Egg Products 2019). This finding also accords with Lusk (2018), who found that the majority of US consumers are WTP a 30+ cents/doz premium for cage-free eggs.

5.2. Real choice experiment estimates

Table 2 reports estimates from the MXL, along with the conditional marginal WTP estimates derived using the conditional inference procedure described in Train (2009) and employed by Lusk and Schroeder (2006).

The estimation results show negative coefficients for both the price and the alternative specific constant representing the no-buy option, suggesting that higher prices are associated with a lower likelihood of purchasing eggs and that individuals preferred to select one of the two experimentally designed alternatives to having nothing at all. Furthermore, we found that consumers prefer quality-differentiated eggs over conventional eggs, as evidenced by the positive and statistically significant coefficients for the USDA organic and cage free labels. Heterogeneity in preferences was observed, with significant standard deviations for both labels.

Looking at the individual specific *m*WTP, the results suggest that consumers are willing to pay the highest price premium for USDA organic eggs ($0.727)

16 To assess the robustness of our results, we estimated three segmented Tobit models, one for each EA. The results are consistent with those produced by the pooled Tobit models. In addition, following Burke (2009), we estimated a Double-Hurdle (DH) model and calculated the partial effect of each EA on three aspects: (a) the probability that WTP > 0, the expected value of WTP given WTP > 0 and the unconditional expected value of WTP. The findings from the first hurdle indicate probabilities of 74 per cent, 88 per cent and 77 per cent for positive bids (bid > 0) in the SPA, BDM and RNPA auctions, respectively. The outcomes from the second hurdle indicate that the expected value of actual bids, conditional on the bids being positive, is $1.62, $1.84 and $1.60 for the SPA, BDM and NPA, respectively. Whereas the 'unconditional' expected value of WTP is $1.20, $1.61 and $1.23 for the SPA, BDM and RNPA, respectively. Overall, our DH analysis aligns with the results of the Tobit model and provides further evidence that the BDM auction yields higher WTP estimates than the SPA and RNPA auctions.

Table 2. Estimates for the mixed logit model in preference space

			Model statistics	
Log likelihood			−163.438	
Choices			744	
Number of parameters			16	
Bayesian Information Criterion			432.7	
Variable	Coefficient	Estimates	Standard errors	|*T*-Statistic|
Price	μ	−1.983***	0.284	6.97
USDA organic	μ	1.435***	0.503	2.85
	σ	1.613***	0.631	2.56
Cage free	μ	0.741*	0.414	1.79
	σ	1.188***	0.476	2.48
No-buy	μ	−4.458***	1.013	4.40
	σ	4.460***	0.993	4.49
Conditional (individual-specific) marginal WTP estimates[b]				
USDA organic	μ	0.727***	0.062	11.760
Cage free	μ	0.376***	0.046	8.217

Note: Single asterisk (*), double asterisk (**) and triple asterisk (***) denote the significance level at 10 per cent, 5 per cent and 1 per cent, respectively. [a]The conditional (individual-specific) WTP for the USDA and cage free labels were computed following the methods described in Train (2009: 259–281).

followed by cage-free eggs ($0.376),[17] which is in line with previous studies (Lusk, 2018) and the results from the EAs.

5.3. Comparing estimates across real choice and experimental auctions

We first present the results from the descriptive statistics and hypothesis tests (see Table A5 Appendix A, supplementary data at ERAE online). The comparison of RCE and EAs is based on mean *m*WTP values[18] disaggregated by experiment (SPA, RNPA, BDM and RCE) and egg type (USDA organic and cage free). The results from the hypothesis tests indicate that the *m*WTP for cage-free is statistically significantly higher in the BDM treatment ($0.662) compared to the SPA ($0.281), RNPA ($0.388) and RCE ($0.376) experiments. The BDM also reveals significantly higher *m*WTP for the USDA-organic label ($0.978) compared to the other experiments ($0.725, $0.682 and $0.727 in the SPA, RNPA and RCE, respectively). However, evidence for statistical significance for BDM yielding higher mWTPs is slightly weaker in the USDA organic comparison. We now present the results from the three Tobit

17 The conditional (individual-specific) marginal WTP mean estimates for the cage free and USDA organic are almost identical to the unconditional (population) marginal WTP mean estimates (0.724 and 0.374 for the USDA organic and cage free eggs, respectively). This supports the proof of equivalence illustrated in Hensher, Rose and Greene (2015), chapter 8, Section 8.1.

18 As mentioned in section 4.3, the *m*WTPs for each product type are calculated as the difference between actual bids for USDA organic/cage free eggs and conventional eggs in the EAs, while for the RCE we employed the inference procedure described in Train (2009).

Table 3. Comparisons of marginal WTP across experiments, estimates from the Tobit/MXL models and hypothesis tests

	Segmented Tobit models			MXL model
Experiments	SPA	BDM	RNPA	RCE
Marginal WTP				
USDA organic vs. conventional	0.863*** (0.148)	1.035*** (0.115)	0.812*** (0.127)	0.729*** (0.062)
Cage free vs. conventional	0.351** (0.149)	0.712*** (0.115)	0.464*** (0.127)	0.373* (0.046)
Hypothesis test and P-values from Poe, Giraud and Loomis (2005) test				
USDA organic				
SPA—BDM = 0	0.164			
SPA—RNPA = *0*	0.400			
BDM—RNPA = *0*	0.085			
RCE—SPA = 0	0.320			
RCE—RNPA =*0*	0.379			
RCE—BDM = 0	0.125			
Cage-free				
SPA—BDM = 0	0.026			
SPA—RNPA = *0*	0.279			
BDM—RNPA = *0*	0.074			
RCE—SPA = 0	0.477			
RCE—RNPA =*0*	0.348			
RCE—BDM = 0	0.080			

Note: Numbers in parenthesis are standard errors. Single asterisk (*), double asterisk (**) and triple asterisk (***) denote the significance level at 10 per cent, 5 per cent and 1 per cent, respectively.

models,[19] one for each EA mechanism, along with the results from the MXL model for the RCE data (see Table 3). The table also reports the p-values from a Poe et al.'s (Poe *et al.*, 2005) test, which was conducted to test for differences in *m*WTP distributions between experiments.[20]

The results from the models confirm the descriptive statistics for the Cage-Free eggs, with less pronounced differences in *m*WTP for USDA organic eggs across experiments. According to the Poe test, the *m*WTP for USDA organic eggs is equivalent across experiments, except for the BDM, which yields statistically higher *m*WTPs compared to RNPA. Conversely, *m*WTPs for Cage-Free eggs is significantly higher in the BDM experiment when compared to all other experiments, while no statistically significant differences in *m*WTP for Cage-Free eggs are observed across the RCE, SPA and RNPA experiments.

19 We report the coefficients from the Tobit model instead of marginal effects, as they can be interpreted as the mean of the uncensored distribution (i.e. the distribution that theoretically allows negative bids) (Canavari *et al.*, 2019), making the RCE and EAs comparable.

20 As noted by a reviewer, the SPA and RNPA auctions might allow for session-based social dynamics not seen in BDM or RCE formats. To check for potential session effects, fixed Tobit models with errors clustered at the session level were also estimated for the SPA and RNPA experiments. The results from these models are consistent with those reported in Table 3.

This evidence aligns with Gracia, Loureiro and Nayga (2011), who found that differences in WTP estimates across RCE and RNPA can depend on the product under consideration but contradicts the findings from Lusk and Schroeder (2006), Shi, Xie and Gao (2018) and Cerroni *et al*. (2019), who found that the RCE generates significantly higher WTP estimates than the SPA. One possible explanation for the divergence between our results and earlier findings could be due to differences in the elicitation methods employed. We conducted the RCE using only four choice tasks, and the EAs involved only three bids (one for each product). This was done to maintain consistency in the elicitation methods, such as time, workload and incentives. Lusk and Schroeder (2006), Gracia, Loureiro and Nayga (2011) and Cerroni *et al*. (2019) utilized 5 bidding rounds and 17 choice tasks, 1 bidding round and 16 choice tasks and 9 bidding rounds and 9 choice tasks, respectively.

6. Robustness check: induced value experiments

Past studies within the agricultural and food space have utilized IV experiments to test the demand-revealing nature of EA mechanisms and RCEs (Lusk and Rousu, 2006; Cerroni *et al*., 2019). Other researchers have used induced value experiments to examine the performance of the econometric models that underlie various empirical assumptions regarding WTP for certain goods (Bazzani, Palma and Nayga, 2018) and to address hypothetical bias in choice modelling, particularly in the context of food choices (Luchini and Watson, 2014; Chavez *et al*., 2020). We use IVEs to provide more robust conclusions about varying bid/choice behaviours across RCE and EAs. In what follows, we describe the experiments, estimation methods and corresponding results of the IVEs.

6.1. Induced value experiments

We conducted four IV experiments, each corresponding to one of the mechanisms studied: SPA-IV, RNPA-IV, BDM-IV and RCE-IV. The experiments were conducted with the same subject pool (consumers) and within the same timeframe as the homegrown versions. In total, 264 subjects participated, with 60, 67, 69 and 68 individuals taking part in the SPA-IV, BDM-IV, RNPA-IV and RCE-IV experiments, respectively. Table A6 in Appendix A (supplementary data at ERAE online) reports descriptive statistics of our sample.

To ensure consistency with our homegrown experiments, we based the IV-EAs on those outlined in previous studies, such as Banerji and Gupta (2014) and Drichoutis, Lusk and Nayga (2015). Similarly, we designed the RCE-IV experiment to match the features of our homegrown RCE, drawing inspiration from Bazzani, Palma and Nayga (2018) and Luchini and Watson (2014). Accordingly, in the EA-IVs, participants placed bids for four tokens that varied in colour (red and blue) and shape (triangle and square). On the other hand, in the RCE-IV, participants were presented with four choice tasks, and they were asked to make a choice in each choice task between two tokens and the option labelled as ‘none of these’. The tokens were described by three attributes that varied at different levels: colour (red and blue), shape (triangle and square) and price ($1.59, $2.59, $3.59 and $4.59). The allocation of attributes and attribute levels was designed using Street and Burgess (2007). Similar to our homegrown experiments, this resulted in eight choice tasks that were randomly grouped into two blocks of four questions each, achieving a D-efficiency of 96.6 per cent.

Prior to the start of each IV experiment, participants were provided with detailed instructions and a table that described the values of the levels of the colour and shape attributes. They were made fully aware of the value of each level and were permitted to consult this table throughout the entirety of the experiment. By ensuring that our participants had access to this information, we aimed to provide them with all the necessary details and resources to make informed decisions. The instructions for the IV-EAs and IV-RCE experiments are available at ERAE online.

6.2. Data analysis and results

In IV-EAs, subjects maximize their payoff by bidding an amount of money equal to the cost or resale value of the token, which is also referred to as the induced value (Noussair, Robin and Ruffieux, 2004; Lusk and Rousu, 2006; Cerroni *et al.*, 2019). In IV-RCE, maximum payoff behavior can be determined through several methods: (1) by comparing the WTP for the token to its induced value (Collins and Vossler, 2009; Bazzani, Palma and Nayga, 2018; Chavez *et al.*, 2020); (2) by examining whether the average marginal utility of the token (selling value) equals the negative value of the average marginal disutility of the price (buying price) (Cerroni *et al.*, 2019); and (3) by calculating the maximum payoff alternative (MPA) for each choice task, where the MPA is defined as the choice alternative with the highest induced value in each task (Luchini and Watson, 2014; Cerroni *et al.*, 2019).

In this application, we compare the demand revelation performance of the four IVEs by examining the individual profit-maximizing behaviour in each experiment, namely the bids equal to the IVs (Bid = IV) for the IV-EAs and MPA for the IV-RCE. However, it is important to note that the selection of the IV may happen with a higher probability in the case of the RCE in comparison with the IV-EAs. This is because in the RCE, if all the participants choose one of the three alternatives randomly, we would still observe nearly a 33 per cent rate of ‘correct’ MPA. In contrast, in the IV-EAs, there is a lower chance of selecting a bid that is exactly equal to the IV if the participant makes a random choice. To balance the probability of choosing MPA relative to BID = IV across experiments, we defined ranges centred at the IVs, with a nearly 33 per cent probability of being randomly selected[21] ($1–$3 for the $2 IV round, $2–$4 for the $3 IV round, $3–$5 for the $4 IV round and $4–$6 for the $5 IV round). Specifically, for the IV-EAs, we used (1) bids equal to

21 The range of all possible bids was between $0 and $6. We used the ± $1 range around each of the induced values, which represent the 33 per cent of the possible value distribution.

the IVs (Bid = IV); (2) bids included in the 33 per cent probability range (BID =33 per cent IV); (3) bids equal to IVs in all four rounds in the EA-IVs (ALLBID = IV); (4) the percentage of individuals with bids included in the 33 per cent probability range in all four rounds in the EA-IVs (ALLBID = 33 per cent IV). For the RCE, we used: (i) the choice tasks where the chosen alternative is the MPA and (ii) individuals having chosen the MPA in all the choice tasks in (ALLMPA). Descriptive statistics for these measures can be found in Table A7 in Appendix A (supplementary data at ERAE online).

Our results show that the percentage of demand-revealing choices in the IV-RCE is higher than any of the IV-EAs, with 56 per cent for ALLMPA and 81 per cent for MPA. This finding is consistent with Cerroni *et al.* (2019) and Luchini and Watson (2014). When considering just the IV-EA mechanisms, the best performing EA is the IV-SPA, albeit it still performs more poorly than the IV-RCE even when using the looser measures that allow for the 33 per cent probability range. The IV-BDM outperforms IV-RNPA when using the looser 33 per cent probability range measures, but it performs worse when considering stricter measures (bids = IV) (5 per cent and 1 per cent vs 10 per cent and 6 per cent for IV-RNPA). In light of these findings, it is reasonable to conclude that the SPA-IV is the most effective mechanism among the EA approaches that we evaluated.

These findings are further validated by our sequential comparative analysis,[22] which was based on the estimation of two Poisson models, whose results are presented in Table 4. The first Poisson model has the number of bids equal to the induced value for the IV-EAs data and the number of chosen MPA for the RCE as the dependent variable (Bid = IV or MPA). The second Poisson model has the number of bids within the 33 per cent probability range for the IV-EA data and the number of chosen MPA in the RCE as the dependent variable (BID = 33 per cent IV or MPA). In both models, SPA is treated as baseline.

Moreover, following previous studies (e.g. McCallum *et al.*, 2022; Fochmann *et al.*, 2021), we computed Wald test statistics for the joint hypothesis of equality between the estimated IV experiments' mean coefficients. The Wald tests conducted in Model 1 (Bid = IV or MPA) indicate that the hypothesis of equality is always rejected at the 1 per cent significance level when comparing IV-RCE estimates with any of the IV-EA experiments. This suggests that the RCE mechanism is more likely to induce the revelation of the IV than any of the EA mechanisms. Indeed, Model 1 suggests that bid equals the induced value 190.1 per cent more frequently with RCE compared to SPA.

In contrast, no significant differences were found between the IV-SPA and IV-RNPA experiments, which outperformed the BDM mechanism. For

22 We also conducted analyses to examine the demand-revealing nature of each mechanism in absolute terms. For the EAs, we utilized the joint Wald test proposed by Shogren *et al.* (2001), whereas for the RCE we employed the hypothesis test proposed by Cerroni *et al.* (2019). The results from the joint Wald tests indicate that none of the EA mechanisms are demand revealing. Similar findings are documented in studies by Shogren *et al.* (2001) and Cerroni *et al.* (2019). On the other hand, the RCE is found to be demand revealing, consistent with findings from Bazzani, Palma and Nayga (2018) and Cerroni *et al.* (2019). The results of this additional analysis are available upon request.

Table 4. Results from the Poisson regression, IV experiments

	Poisson regression models	
	Bid = IV + MPA	BID = 33%IV + MPA
BDM	−0.913** (0.449)	−0.058 (0.094)
RNPA	−0.175 (0.392)	−0.228** (0.107)
RCE	1.906*** (0.264)	0.161** (0.078)
Constant	−0.727*** (0.261)	1.018*** (0.068)
Wald Tests[a]		
BDM = RNPA = RCE = 0	<0.001	<0.001
RCE—BDM = 0	<0.001	<0.004
RCE—RNPA = 0	<0.001	<0.001
BDM—RNPA = 0	0.115	0.103
LPL	−281.636	−472.437
No. of observations	264	264

instance, compared to SPA, BDM decreases the frequency of bid equalling the induced value by 91.3 per cent. The results from Model 2 (Bid = 33 per cent IV or MPA) confirm the findings from Model 1, indicating that the RCE mechanism is the most effective in revealing demand: compared to SPA, RCE increases the frequency of bid equalling the induced value by 16.1 per cent. No significant differences are found between the SPA and BDM mechanisms. Compared to SPA, RNPA decreases the frequency of bid equalling the induced value by 22.8 per cent.

7. Discussion and conclusions

This study uses homegrown experiments to explore whether and how valuations from RCEs differ from those derived from three different EAs commonly used in food choice literature: SPA, BDM and RNPA. In addition, to test the robustness of our results, for each homegrown experiment, we also perform equivalent IV experiments.

For the homegrown experiments, we find that the WTP values derived from the RCE are not statistically different from those derived from the SPA and the RNPA when holding the product type (egg-type) constant. We also find that the WTPs derived from the BDM mechanism are higher than the WTPs derived from the other elicitation methods (SPA, RNPA and RCE) while the results from the IV experiments generally suggest that the BDM mechanism is less accurate at revealing bids/choices consistent with underlying induced values. This evidence provides further support for the hypothesis that the BDM mechanism may not be incentive compatible and, thus, not the best choice when it comes to non-market valuation methods (Horowitz, 2006; Canavari *et al.*, 2019). One potential explanation for this result may be the fact that subjects are less familiar with the BDM mechanism, and this might create a barrier in their efforts to reveal their true preferences (Canavari *et al.*, 2019). Importantly, our findings corroborate recent studies that have questioned the incentive compatibility nature of BDM auctions. Psychological factors greatly influence the way people behave in such auctions, as highlighted by Vassilopoulos, Drichoutis and Nayga (2018) and Canavari *et al*. (2019) in their exhaustive and detailed discussions.

BDM auctions are commonly used in agricultural economics, particularly in international development contexts (De Groote, Kimenju and Morawetz, 2011), due to their logistical advantages. RCEs have many of the same logistical advantages as the BDM mechanism (no need to form groups of people to run the experiments; individual decision-making), and their use in logistically difficult situations (e.g. in-store or real points of purchase market research or WTP studies in developing country contexts) may be a good solution. While some researchers may argue that in RCEs, WTP measures are derived from econometric models that require assumptions about the distribution of random coefficients, this approach can lead to more robust welfare estimates despite its potential to complicate estimation procedures. Our homegrown experiments found no differences in welfare estimates across SPA, RNPA and RCE. Consistent with recent research (Cerroni *et al*., 2019), we also found that RCE performs best in IV experiments, further attesting to its demand-revealing nature. Notably, recent developments in the choice modelling literature offer more flexible (semi-parametric) distributions for discrete choice analyses, especially when sample sizes are particularly large (see, for example, Caputo *et al*., 2018; Bazzani, Palma and Nayga, 2018; Scarpa, Franceschinis and Thiene, 2020).

As in other EA applications in the agriculture and food domain, our study uses a limited number of respondents per experiment.[23] We note (as have several recent literature review studies, including Canavari *et al*., 2019; Palm-Forster and Messer, 2021; Caputo and Scarpa, 2022) that although implementation of power calculations prior to data collection in non-market valuation studies of this type is currently limited, future research should also look into this issue. Furthermore, our study employed unpaid practice rounds to familiarize participants with each mechanism. In future studies, it would be beneficial to explore whether the use of different training protocols and paid/unpaid practice rounds could mitigate differences in bidding/choice behaviour across EA mechanisms. Although research in this domain exists (Plott and Zeiler, 2005; Drichoutis, Nayga and Lazaridis, 2011; Cason and Plott, 2014; Corrigan, Rousu and Depositario, 2014; Drichoutis and Nayga, 2022), findings remain scarce and are often inconclusive.[24] In terms of our

23 We conducted a search on the Web of Science to identify studies that have utilized experimental auctions in the food domain, resulting in 66 relevant studies. Out of these studies, we found that 34 per cent had sample sizes of <50 responses, 43 per cent had sample sizes between 50 and 100 participants, and only 23 per cent had sample sizes >100 participants.

24 Drichoutis, Nayga and Lazaridis (2011) looked at the role of training in experimental auctions. However, the treatments implemented by the authors do not directly test for paid versus unpaid practice rounds. Indeed, the authors executed two treatments: extensive training with paid practice round treatment and minimal training with unpaid practice round treatment. Results from this study indicate difference in WTP estimates from the two treatments; bid values of subjects given

experimental design, we focused on accounting for main effects. However, there is a need for additional research that compares welfare estimates between RCE and EAs, considering both main and interaction effects. Such a comparison holds the potential to yield valuable insights, as also discussed by Lusk (2003).

Further work needs to be done to also investigate whether the use of single versus multiple rounds and different price ranges would influence welfare estimates in BDM and RCE. In the case of the SPA and RNPA, it is also worthwhile to investigate the number of bidders within auction groups, with particular attention to whether and how different group sizes influence bidding/choice behaviour and ultimately welfare estimates.

Finally, it is important to point out that one explanation behind the more accurate results obtained in the RCE-IV over the EA-IV treatments might be the less effortful mechanism in choosing the maximum payoff alternative in comparison with selecting the induced value. Emerging literature shows that individuals with higher cognitive ability tend to behave in closer accordance with theory (Lee *et al.*, 2020). Hence, we recommend that future research test whether cognitive ability is a significant factor in shaping bid/choice behaviour across different preference elicitation mechanisms.

Acknowledgements

The authors would like to express their gratitude to the editor, Carl-Johan Lagerkvist and three anonymous reviewers for their helpful comments and suggestions. In addition, the authors extend their thanks to Andreas Drichoutis and Simone Cerroni for their valuable feedback. This work was partly funded by the USDA National Institute of Food and Agriculture, Hatch Project No. 1013332, and the National Science Foundation DGE-1828942.

Supplementary data

Supplementary data are available at *ERAE* online.

Conflicts of interest

The authors have nothing to declare here.

extensive training and paid practice round were higher than those of subjects given only minimal training prior to the actual auction and unpaid practice round. Yet, it is still unclear whether this difference is driven by the extensive training or by the paid/unpaid practice round as their design does not allow to isolate these effects (training versus paid/unpaid practice round). Plott and Zeiler (2005) discuss the importance of training and practice rounds before actual BDM experiments are conducted, but again in their study (which mused on the WTP/WTA gap), they compare paid practice rounds to no practice rounds, so the issue of paid/unpaid is not addressed directly. Corrigan, Rousu and Depositario (2014) study whether practice rounds impact auction results but they do not directly address paid versus unpaid practice rounds.

References

Alfnes, F. (2009) Valuing product attributes in Vickrey auctions when market substitutes are available. *European Review of Agricultural Economics* 36: 133–149.

Alfnes, F., Guttormsen, A. G., Steine, G. and 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: 1050–1061.

Banerji, A. and Gupta, N. (2014) Detection, identification, and estimation of loss aversion: evidence from an auction experiment. *American Economic Journal: Microeconomics* 6: 91–133.

Bazzani, C., Caputo, V., Nayga, R. M., Jr and 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.

Bazzani, C., Palma, M. A. and Nayga, R. M., Jr (2018) On the use of flexible mixing distributions in WTP space: an induced value choice experiment. *Australian Journal of Agricultural and Resource Economics* 62: 185–198.

Becker, G. M., DeGroot, M. H. and Marschak, J. (1964) Measuring utility by a single-response sequential method. *Behavioral Science* 9: 226–232.

Brooks, K. and Lusk, J. L. (2010) Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *American Journal of Agricultural Economics* 92: 1229–1241.

Burke, W. J. (2009) Fitting and interpreting Cragg's Tobit alternative using Stata. *The Stata Journal* 9: 584–592.

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

Caputo, V. and Scarpa, R. (2022) Methodological advances in food choice experiments and modeling: current practices, challenges, and future research directions. *Annual Review of Resource Economics* 14: 63–90.

Caputo, V., Scarpa, R., Nayga, R. M. and Ortega, D. L. (2018) Are preferences for food quality attributes really normally distributed? An analysis using flexible mixing distributions. *Journal of Choice Modeling* 28: 10–27.

Cason, T. N. and Plott, C. R. (2014) Misconceptions and game form recognition: challenges to theories of revealed preference and framing. *Journal of Political Economy* 122: 1235–1270.

Cerroni, S., Watson, V., Kalentakis, D. and Macdiarmid, J. I. (2019) Value-elicitation and value-formation properties of discrete choice experiment and experimental auctions. *European Review of Agricultural Economics* 46: 3–27.

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

Charness, G., Gneezy, U. and Kuhn, M. A. (2012) Experimental methods: between-subject and within-subject design. *Journal of Economic Behavior & Organization* 81: 1–8.

Chavez, D. E., Palma, M. A., Nayga, R. M., Jr and Mjelde, J. W. (2020) Product availability in discrete choice experiments with private goods. *Journal of Choice Modelling* 36: 100225.

Collins, J. P. and Vossler, C. A. (2009) Incentive compatibility tests of choice experiment value elicitation questions. *Journal of Environmental Economics and Management* 58: 226–235.

Corrigan, J. R., Rousu, M. C. and Depositario, D. P. T. (2014) Do practice rounds affect experimental auction results? *Economics Letters* 123: 42–44.

De Groote, H., Kimenju, S. C. and Morawetz, U. B. (2011) Estimating consumer willingness to pay for food quality with experimental auctions: the case of yellow versus fortified maize meal in Kenya. *Agricultural Economics* 42: 1–16.

Demont, M., Zossou, E., Rutsaert, P., Ndour, M., Van Mele, P. and Verbeke, W. (2012) Consumer valuation of improved rice parboiling technologies in Benin. *Food Quality and Preference* 23: 63–70.

Drichoutis, A. C., Lusk, J. L. and Nayga, R. M. (2015) The veil of experimental currency units in second-price auctions. *Journal of the Economic Science Association* 1: 182–196.

Drichoutis, A. C. and Nayga, R. M., Jr (2022) Game form recognition in preference elicitation, cognitive abilities, and cognitive load. *Journal of Economic Behavior & Organization* 193: 49–65.

Drichoutis, A. C., Nayga, R. M. and Lazaridis, P. (2011) The role of training in experimental auctions. *American Journal of Agricultural Economics* 93: 521–527.

Fochmann, M., Fochmann, N., Kocher, M. G. and Müller, N. (2021) Dishonesty and risk-taking: compliance decisions of individuals and groups. *Journal of Economic Behavior & Organization* 185: 250–286.

Gracia, A., Loureiro, M. L. and Nayga, R. M., Jr (2011) Are valuations from nonhypothetical choice experiments different from those of experimental auctions? *American Journal of Agricultural Economics* 93: 1358–1373.

Grebitus, C., Lusk, J. L. and Nayga, R. M., Jr (2013) Explaining differences in real and hypothetical experimental auctions and choice experiments with personality. *Journal of Economic Psychology* 36: 11–26.

Haghani, M., Bliemer, M. C., Rose, J. M., Oppewal, H. and Lancsar, E. (2021) 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* 41: 100309.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2015) Applied choice analysis.

Horowitz, J. K. (2006) The Becker–DeGroot–Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letter* 93: 6–11.

Karni, E. and Safra, Z. (1987) "Preference reversal" and the observability of preferences by experimental methods. *Econometrica: Journal of the Econometric Society* 55: 675–685.

Krinsky, I. and Robb, A. L. (1986) On approximating the statistical properties of elasticities. *The Review of Economics and Statistics* 68: 715–719.

Lee, J. Y., Nayga, R. M., Jr, Deck, C. and Drichoutis, A. C. (2020) Cognitive ability and bidding behavior in second price auctions: an experimental study. *American Journal of Agricultural Economics* 102: 1494–1510.

List, J. A. (2003) Using random nth price auctions to value non-market goods and services. *Journal of Regulatory Economics* 23: 193–205.

Lonati, S., Quiroga, B. F., Zehnder, C. and Antonakis, J. (2018) On doing relevant and rigorous experiments: review and recommendations. *Journal of Operations Management* 64: 19–40.

Louviere, J. J. and Islam, T. (2008) A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research* 61: 903–911.

Luchini, S. and Watson, V. (2014) Are choice experiments reliable? Evidence from the Lab. *Economics Letters* 124: 9–13.

Lucking-Reiley, D. (1999) Using field experiments to test equivalence between auction formats: magic on the internet. *American Economic Review* 89: 1063–1080.

Lusk, J. L. (2003) Using experimental auctions for marketing applications: a discussion. *Journal of Agricultural and Applied Economics* 35(2): 349–360.
Lusk, J. L. (2018) Market potential for cage free eggs. http://jaysonlusk.com/blog/2018/1/1/market-potential-for-cage-free-eggs. Accessed February 21, 2019.
Lusk, J. L., Feldkamp, T. and Schroeder, T. C. (2004) Experimental auction procedure: impact on valuation of quality differentiated goods. *American Journal of Agricultural Economics* 86: 389–405.
Lusk, J. L. and Rousu, M. (2006) Market price endogeneity and accuracy of value elicitation mechanisms. In: List J. A. (ed.), *Using Experimental Methods in Environmental and Resource Economics*. Cheltenham, United Kingdom: Edward Elgar Publishing, Cheltenham, 20–40.
Lusk, J. L. and Schroeder, T. C. (2004) Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics* 86: 467–482.
Lusk, J. L. and Schroeder, T. C. (2006) Auction bids and shopping choices. *Advances in Economic Analysis & Policy* 6(1).
Lusk, J. L., and Shogren, J. F. (2007). *Experimental Auctions*. Cambridge: J. Econ.
Machina, M. J. (1982) "Expected utility" analysis without the independence axiom. *Econometrica: Journal of the Econometric Society* 50: 277–323.
McCallum, C. S., Cerroni, S., Derbyshire, D., Hutchinson, W. G. and Nayga, R. M., Jr (2022) Consumers' responses to food fraud risks: an economic experiment. *European Review of Agricultural Economics* 49: 942–969.
Murphy, J. J., Stevens, T. H. and Yadav, L. (2010) A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. *Environmental and Resource Economics* 47: 111–123.
Noussair, C., Robin, S. and Ruffieux, B. (2004) Revealing consumers' willingness-to-pay: a comparison of the BDM mechanism and the Vickrey auction. *Journal of Economic Psychology* 25: 725–741.
Palm-Forster, L. H. and Messer, K. D. (2021) Experimental and behavioral economics to inform agri-environmental programs and policies. In: Barrett C. B. and Just D. R. (eds), *Handbook of Agricultural Economics, Vol. 5*, Amsterdam, Netherlands: Elsevier, 4331–4406.
Penn, J. M. and Hu, W. (2018) Understanding hypothetical bias: an enhanced meta-analysis. *American Journal of Agricultural Economics* 100: 1186–1206.
Plott, C. R. and Zeiler, K. (2005) The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review* 95: 530–545.
Poe, G. L., Giraud, K. L. and Loomis, J. B. (2005) Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics* 87: 353–365.
Rosato, A. and Tymula, A. A. (2019) Loss aversion and competition in Vickrey auctions: money ain't no good. *Games and Economic Behavior* 115: 188–208.
Rutström, E. E. (1998) Home-grown values and incentive compatible auction design. *International Journal of Game Theory* 27: 427–441.
Scarpa, R., Franceschinis, C. and Thiene, M. (2020) Logit mixed logit under asymmetry and multimodality of WTP: a Monte Carlo evaluation. *American Journal of Agricultural Economics* 103: 643–662.
Shehata, E. (2011) TOBITHETM: stata module to estimate Tobit multiplicative heteroscedasticity regression.

Shi, L., Xie, J. and Gao, Z. (2018) The impact of deal-proneness on WTP estimates in incentive-aligned value elicitation methods. *Agricultural Economics* 49: 353–362.

Shogren, J. F., Fox, J. A., Hayes, D. J. and Kliebenstein, J. B. (1994) Bid sensitivity and the structure of the Vickrey auction. *American Journal of Agricultural Economics* 76: 1089–1095.

Shogren, J. F., Margolis, M., Koo, C. and List, J. A. (2001) A random nth-price auction. *Journal of Economic Behavior & Organization* 46: 409–421.

Smith, V. L. (1976) Experimental economics: induced value theory. *The American Economic Review* 66: 274–279.

Street, D. J. and Burgess, L. (2007) *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. Vol. 647. Hoboken, NJ, USA: John Wiley & Sons.

Train, K. E. (2009) *Discrete Choice Methods with Simulation*. United States: Cambridge University Press, United States Code of Federal Regulations, 7 C.F.R. § 205.239. (2002).

USDA national retail report – shell egg and egg products. (2019, November 8). Available from https://www.ams.usda.gov/mnreports/pywretailegg.pdf. Accessed November 11 2019.

Vassilopoulos, A., Drichoutis, A. C. and Nayga, R. M., Jr (2018). Loss aversion, expectations and anchoring in the BDM mechanism. Munich Personal RePEc Archive No. 85635.

Vickrey, W. (1961) Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance* 16: 8–37.