AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Preferences for healthy and environmentally sustainable food: Combining induced-value and home-grown experiments

S. Cerroni[1]; V. Watson[2]; J. Macdiarmid[2]

1: Queen's University Belfast, , United Kingdom, 2: University of Aberdeen, , United Kingdom

Corresponding author email: s.cerroni@qub.ac.uk

Abstract:

This paper tests if Second Price Vickrey Auction (SPVA) and Discrete Choice Experiment (DCE) are isomorphic and whether lack of isomorphism is due to value-elicitation, value-formation or both. We conduct an artefactual field experiment that combines induced-value (IV) and home-grown (HG) procedures using SPVA and DCE. IV preferences are elicited for tokens and HG preferences for multi-attribute lasagnes. Attributes are healthiness and environmental sustainability. Our results suggest that HG preferences differ across elicitation methods. This discrepancy is due to value-elicitation and value-formation. DCE is the most demand-revealing approach and provides the highest premiums for healthy and environmentally sustainable lasagnes.

Acknowledegment:

JEL Codes: D12, Q18

#1139

Preferences for healthy and environmentally sustainable food: Combining induced-value and home-grown experiments

Abstract

This paper tests if Second Price Vickrey Auction (SPVA) and Discrete Choice Experiment (DCE) are isomorphic and whether lack of isomorphism is due to value-elicitation, value-formation or both. We conduct an artefactual field experiment that combines induced-value (IV) and home-grown (HG) procedures using SPVA and DCE. IV preferences are elicited for tokens and HG preferences for multi-attribute lasagnes. Attributes are healthiness and environmental sustainability. Our results suggest that HG preferences differ across elicitation methods. This discrepancy is due to value-elicitation and value-formation. DCE is the most demand-revealing approach and provides the highest premiums for healthy and environmentally sustainable lasagnes.

1. Introduction

Food economists have a long-standing interest in consumers' preferences for new products and food policies, such as food labelling. These preferences are needed either to gauge the demand for new products or to evaluate a food policy's potential effectiveness before it is introduced. In both of these cases, market data are not yet available, and, hence, entrepreneurs and policy makers need robust techniques that can reliably elicit home-grown (HG) preferences for food. Home-grown preferences are those which are genuinely formed by people without any direct interference from researchers about the value of the good under study (Rutström, 1998).

The techniques that are most commonly used to elicit HG preferences are discrete choice experiments (DCEs), which are based on based on Lancaster's theory of value (Lancaster, 1966) and experimental auctions (EAs). The second-price Vickrey auction (SPVA) (Vickrey, 1961) is one of the most frequently applied EAs because it can be easily implemented in the lab as well as in the field, and laypeople appear to understand the mechanism better than they understand other EAs. Both DCEs and SPVAs are *theoretically* demand revealing, which means that under a correct monetary incentive scheme these techniques should induce subjects to truthfully reveal their real preferences (Cummings *et al.*, 1997). Therefore, preferences elicited in DCEs and SPVA should be isomorphic. In other words, the HG preference order elicited should be invariant to the elicitation technique (Coppinger, Smith and Titus, 1980).

Yet studies find that HG preferences for food items vary with the elicitation mechanism used (e.g., Lusk and Schroeder, 2006; Gracia, Loureiro and Nayga, 2011). There are two explanations for these empirical results. First, some elicitation mechanisms are not demand revealing *in practice* (value-elicitation problem). Second, there are differences in the way HG preferences are formed across elicitation mechanisms (value-formation problem)

(Lusk and Schroeder, 2006). Studies that only elicit HG preferences for food items cannot distinguish between these two explanations.

Induced-value (IV) experiments provide a method that researchers can use to test if an elicitation mechanism is demand revealing *in practice* and, therefore IV experiments can detect value-elicitation problems (Vossler and McKee, 2006). In IV experiments, subjects are asked to express their willingness to pay (WTP) for fictitious goods (for example, tokens) whose values are given in the experiment. The induced value is the price the good can be exchanged for in the experiment (Smith, 1976).

In this study, we test whether a lack of isomorphism across SPVAs and DCEs is due to value-elicitation problems by using an artefactual field experiment (Harrison and List, 2004) that elicits both IV and HG preferences for private goods. We use a between-sample design and assign subjects to one of two treatment groups: SPVA or DCE. In each group, subjects complete both IV and HG elicitation tasks. The IV tasks measure demand revelation for a token, while the HG tasks elicit preferences for a real food item, a beef-based lasagne that is labelled using a traffic light system (TLS) for both environmental sustainability and healthiness. This study aims to identify the most demand-revealing elicitation approach between SPVA and DCEs, and, hence estimate the "truer" preferences for healthy and environmentally sustainable beef lasagne. Additionally, we investigate whether value formation problems contribute to the lack of isomorphism across SPVAs and DCEs.

Our results suggest that the HG preferences elicited in the SPVA and DCE tasks are not isomorphic. We show that differences in HG preferences are caused by both value-elicitation and value-formation differences. The DCE elicits higher WTP estimates for our lasagnes than the SPVA. This result implies that using mechanisms that are not demand revealing can underestimate the premia that consumers are willing to pay for food products.

This study contributes to the existing literature in two ways. First, this is the first study to explain the lack of isomorphism across elicitation mechanism by combining IV and HG procedures. Second, this is the first study to elicit consumers' preferences for both healthy and environmentally sustainable products. Previous studies have elicited consumers' preferences for either healthy or environmentally sustainable products, but have not combined both dimensions (e.g., Belcombe, Fraser and Di Falco, 2010; Caputo, Nayga and Scarpa 2013; Akaichi, Nayga and Nalley, 2017).

2. The literature on demand-revelation

2.1. *Demand-revelation in the SPVA*

Although the SPVA have been regarded as theoretically demand-revealing, empirical results about demand revelation are mixed. Early IV investigations confirm demand revelation in SPVAs (e.g., Coppinger, Smith and Titus, 1980; Cox, Roberson and Smith, 1982), however later studies consistently find that subjects tend to overbid (e.g., Kaegel, Harstad and Levin, 1987; Kaegel and Levine, 1993). Recent studies show that SPVA performs well at the aggregate level, but generates extreme behaviour at the individual level (e.g., Shogren *et al.*, 2001; Noussair, Robin and Ruffieux, 2004; Lusk and Rousu, 2006). These studies find that subjects tend to underbid with respect to the dominant strategy.

Horowitz (2006) proved that these mixed results occur because mechanisms are not always theoretically incentive compatible. Building on Karni and Safra's (1987) work on the incentive compatibility of the Becker deGroot Marschak (BDM) (Becker, DeGroot and Marschak, 1964) mechanism in which lotteries are at stake, he proved that BDM and SPVA are not incentive compatible when subjects bid according to some forms of non-expected utility theory. To date, no empirical study has tested whether the SPVA fails to be demand revealing because of the Horowitz theory or because details of the experimental design lead

subjects to fail to identify the optimal strategy for example due to the low cost of deviations from the optimal strategy.

2.2. Demand-revelation in the DCE

The theoretical incentive compatibility of hypothetical DCE responses about public goods is a controversial topic that has received much attention in the stated-preference literature (see discussions in Cummings *et al.*, 1997; Harrison, 2006; Carson and Groves, 2007). In a private good context, DCEs should replicate the results of revealed preference analyses if individuals face the real consequences of their choices (i.e., a transaction takes place), only one choice situation determines real consequences, and they fully understand instructions.

Two IV studies test demand revelation in DCEs. Collins and Vossler (2009) use referenda-style DCEs to elicit IV preferences for public goods and services, and Luchini and Watson (2014) elicit IV preferences for a private good. The studies provide mixed results: Collins and Vossler (2009) find a high level of demand revelation and Luchini and Watson (2014) a moderate level. However, the cognitive demands of the DCE tasks differ across the studies: Collins and Vossler (2009) elicit preferences for three attribute bundles in which each attribute is numerical using six choice sets, whereas Luchini and Watson (2014) elicit preferences for four attribute bundles in which three attributes are categorical using nine choice sets.

2.3. Isomorphism in home-grown preferences

Although SPVA and DCE mechanisms should be isomorphic, many studies investigating HG preferences for food products have shown that they are not. Lusk and Schroeder (2006) found SPVA bids were significantly lower than WTP estimated from DCE

when eliciting students' HG preferences for steaks. Other studies have found that valuation differences across mechanisms are correlated with individual's characteristics. Gracia, Loureiro, and Nayga (2011) found that differences between random nth auction bids and DCE WTP estimates depend on individuals' socio-demographic characteristics. Grebitus, Lusk and Nayga (2013) found that discrepancies between HG preferences for apple and wine products elicited using SPVA and DCE are related to subjects' personality traits.

3. Methods

3.1. Case study

The food that we purchase and eat not only affects our health, but also the environment. Diets with high quantities of meat are environmentally unsustainable because of livestock production is associated with the high levels of greenhouse gas emission (GHGE) (e.g., Carlsson-Kanyama and González, 2009). In the UK, many people consume large quantities of meat and they do not associate meat with GHGEs (Macdiarmid, Douglas and Campbell, 2016). This study tests a labelling approach to encourage consumers to reduce their meat intake.

In the UK, consumers are informed about the nutritional properties of products that are sold in the market via food labels. An example is the Traffic Light System (TLS) developed by the Food Standard Agency (FSA, 2013). The TLS labels typically indicate the levels for four key nutrients: fat, saturated fat, sugar and salt. These levels are mapped using a chromatic scale: red for high content, amber for medium content, and green for low content (FSA, 2013). On the other hand, consumers have little guidance about the negative externalities that food production has on the environment. In our study, we develop a new TLS related to carbon footprint where: red stands for high, amber for medium, and green for

low carbon footprint.[1] We then use the new TLS alongside the traditional nutritional information TLS to investigate the extent to which consumers trade-off health and environmental sustainability.

Our study focus on preferences for a specific beef-based food: frozen lasagne. The healthiness and environmental sustainability of this food item can be manipulated by changing the proportions of the traditional ingredients (e.g. beef, pasta, sauce, cheese). We use the saturated fat content of the food (saturated fat g/100g product) as the indicator for the healthiness and the greenhouse gas emissions (GHGE) (kgCO2e/100g product) as an indicator for the environmental sustainability.[2] We maintain a similar appearance and same portion size (400 grams) across the dishes to minimise selection bias based on the visual appeal of the food.

There are nine lasagne options; three levels of healthiness by three levels of GHGE. The healthiness is varied by manipulating the saturated fat content of the lasagne. We used different grades of scotch beef (i.e. fatty/lean), different cheese types (i.e. full fat/reduced fat) and the amount of fat added in the cooking process. The GHGE is varied by using different quantities of beef, cheese and vegetables. The lasagne recipes were designed and prepared at the Human Nutrition Unit (HNU) at The Rowett Institute (University of Aberdeen, United Kingdom). The lasagne were pre-cooked and frozen at the HNU then transported frozen (using cooling bags) to the location of the experiment (SEEL, University of Aberdeen) on the day of the experiments. Subjects were provided with plastic cooling bags to keep any lasagnes they took home frozen.

3.2. Sample and experiment procedures

[1] Carbon footprint represents the greenhouse gas emissions associated with production of a product (Carbon Trust, 2012).

[2] We used the FSA TLS to determine the quantity of saturated fat. Only saturated fat is considered since high intakes has been associated with negative health outcomes and other nutrients are kept constant across alternatives.

Our sample consists of 128 consumers randomly recruited from the population living in or around the city of Aberdeen. The study was advertised as a generic study on consumers' food choices about beef-based products.

The sample for the main study was randomly split across two treatment groups, one for each alternative elicitation mechanism: 63 subjects were allocated to SPVA treatment and 65 to the DCE treatment.[3,4] An in-depth description of each elicitation mechanism is provided in Appendix 1.

All subjects receive a show-up fee of £10 (given to them at the end of the experiment) and perform the following tasks in order: a) complete the IV task after having received instructions on the procedure; b) complete the HG task after having received instructions on the procedure; and finally c) answer a series of debriefing questions about their consumption habits and socio-economic status. Subjects were not informed about the outcome of the IV tasks until the end of the experiment to prevent their earnings in the IV tasks influencing their purchasing behaviour in the HG task. The IV and HG tasks are described in the following section and summarized in Table 1.[5]

Table 1. Steps of Elicitation Procedures

[3] In the SPVA treatment, we ran 8 sessions: four sessions with 8 subjects, two sessions with 9 subjects, one session with 7 subjects and the remaining session with 6 subjects. In the DCE session, we implemented 5 sessions with 10, 9, 18, 9, and 19 subjects, respectively. All DCE sessions and 4 SPVA sessions were run between 28th January 2015 and 24th April 2015. Additional 4 SPVA sessions were run in September 2017. All sessions took place either at 1.30pm or 5.30 pm. This is to avoid any potential confounding effect related to time, hunger and satiety.

[4] Subjects were assigned to treatment groups using the following randomization procedure, each subject has a chance p=0.32 (31 subjects out of 96) to be assigned to the SPVA treatment group and a chance q=0.68 (65 subjects out of 96) to be assigned to the DCE treatment group. We acknowledge that our randomization procedure may be biased because of the additional sampling campaign conducted to run the additional 4 sessions in September 2017.

[5] No practice sessions were run for the IV and HG tasks because the length of experimental instructions and tasks was substantial. However, subjects were extensively instructed about elicitation mechanisms and detailed examples of different choice-behaviour in each treatment were provided. Subjects were also invited to ask questions about tasks, if instructions were not clear before starting their tasks.

a. Induced-value Tasks		
Step	**SPVA**	**DCE**
A	Subjects receive a bidding sheet for 9 numbered tokens, each associated with a resale value	Subjects are presented with 9 choice sets (2 tokens and an opt-out)
B	Subjects bid for each token on the bid sheet	Subject choose the most preferred alternative in each choice set
C	The binding token is randomly drawn	The binding choice set is randomly drawn
D	The winner is identified as the subject who submitted the highest bid	The profit is calculated and paid. Profit = induce value - market price for buyers. Profit = 0 for the others.
E	The market price is identified as the second highest bid	
F	The profit is calculated and paid. Profit = induce value - market price for winners. Profit = 0 for the others	
b. Home-grown Tasks		
Step	**SPVA**	**DCE**
A	Subjects receive a bidding sheet with 9 lasagnes	Subjects are presented with 9 choice sets (2 lasagnes and an opt-out)
B	Subjects bid for each lasagne on the bid sheet	Subject choose the most preferred alternative in each choice set
C	The binding lasagne is randomly drawn	The binding choice set is randomly drawn
D	The winner is identified as the subject who submitted the highest bid	Subjects who buy the lasagne, pay the market price and get the lasagne, while the others do not.
E	The market price is identified as the second highest bid	
F	The winner pays the market price and gets the binding lasagne, while others pay nothing and get nothing.	

3.3. Second-Price Vickrey Auction

In the IV task, subjects are asked to bid on nine tokens numbered from 1 to 9) Induced values for the tokens range from £1.00 to £5.00 in £0.50 increments.[6] Each subject faces each induced value only once and, hence, the whole induced demand curve is induced during the game. Subjects are told that the induced value is the price at which they sell tokens back to the researcher, if they win and buy the token. We call this the resale value of the token. Subjects' earnings depend on their own and others' bids for the one randomly selected token,

[6] As standard in SPVA, subjects were informed that the distribution of IV prices differed across subjects.

the binding token, which is unknown to the subjects prior to bidding. The winner is identified as the subject who submitted the highest bid for the binding token, while the market price at which the winner buys the binding token is identified as the second highest bid.[7] Our experimental instructions follow those provided by Shogren *et al.* (2001).

In the HG task, subjects are asked to bid on each of the nine different lasagnes.[8,9] Subjects may win and buy one lasagne, the binding lasagne.[10] Whether or not a subject wins depends on the subject's own and others' bids. As above, the winner is identified as the subject who submitted the highest bid for the binding lasagne, while the market price at which the winner buys the binding lasagne is identified as the second highest bid. This procedure follows that reported in Lusk and Schroeder (2006) and Gracia, Loureiro and Nayga (2011).

3.4. Discrete Choice Experiment

In the IV task, we mimic a typical DCE and create choice sets that each contain two tokens and an opt-out (i.e. buy neither alternative). Tokens have two attributes: market price and resale value (or induced value). The market price is the price at which subject buys the token, the resale value is the price at which they will sell the token. The IV and market prices range from £1.00 to £5.00 in £0.50 increments. An IV and a market price are allocated to each token using a fractional factorial design (ChoiceMetrics, 2012). In each choice set,

[7] The winner can incur in a loss. In that case, the loss is subtracted from the show-up fee. Subjects are informed about this.

[8] The sequence of lasagne from right to left are randomized. Half of the subjects were presented with lasagnes ordered from red-red to green-green, while the other half of the subjects were presented with an inverted order of lasagne (i.e, from green-green to red-red).

[9] In our instructions, we provide two sequences of bidding examples. Half of them subjects were presented with a sequence in which higher bids were associated with healthy and low carbon footprint lasagne. The other half was presented with a sequence in which higher bids were associated with unhealthy and high carbon footprint lasagne.

[10] To avoid food waste, the binding lasagne was randomly drawn before the experiment was conducted and only portions of the binding lasagne were cooked and carried to the experimental lab. The draw was video-recorded and showed to subjects at the end of the sessions.

subjects are asked to select their most preferred alternative.[11] Each subject's profit is the difference between the induced value and the market price of the chosen alternative in the binding choice set, which is randomly selected (see footnote 7 for subject who incur in losses).

The design of our IV DCE is based on previous studies (Collins and Vossler, 2009; Luchini and Watson, 2014). We focus on a comparison of our study to Luchini and Watson's application (2014), which also elicits IVs for private goods. Luchini and Watson (2014) describe tokens using 4 attributes, each presenting three categorical levels. Colour can be red, blue, or yellow; shape can be circular, squared or triangular; and size can be small, medium, large. Each of these categorical levels is associated to a marginal IV, so that subjects are required to compute the total IV of each token presented in the choice sets by adding marginal IVs. The last attribute is price. The design of our IV task is simpler because we directly present the total IV of each alternative token to subjects, instead of asking subjects to compute this from marginal level-based IVs. Our approach is conceptually equivalent to Luchini and Watson's design and resembles the classic design of IV SPVA studies, where subjects are typically provided with the total IV associated to each fictitious good and not asked to compute it. Our instructions replicate those reported in Collins and Vossler (2009) and Watson and Luchini (2014).

In the HG task, we design a DCE to elicit subjects' preferences for a lasagne described by three attributes: healthiness, carbon footprint, and market price. The healthiness and carbon footprint attributes each have three levels low, medium and high, and are described using a TLS with the colours green, amber and red, respectively. The market price ranges from £1.00 to £5.00 in £0.50 increments. We generate our nine choice sets by using a D-

[11] Choice tasks have been randomized across respondents to control for order effect.

efficient design (ChoiceMetrics, 2012)[12,13]. Each choice task has three alternatives: two lasagne portions and an opt-out (i.e. buy nothing). At the end of the experiment, each subject pays the market price of the chosen lasagne (if any) in the binding choice set and gets the lasagne. Subjects who selected the opt-out, pay nothing and get nothing.[14] Our instructions follow those reported in Lusk and Schroeder (2006) and Gracia, Loureiro and Nayga (2011).

4. Testing isomorphism using HG procedures

4.1. Testable hypotheses and model specifications

We test the consistency in HG preferences across the SPVA and DCEs (i.e., isomorphism) using a two-step procedure (Lusk and Schroeder, 2006; Gracia, Loureiro, and Nayga, 2011). First, models for panel data are estimated using SPVA data and Random Utility Models (RUMs) (McFadden 1973) are estimated using DCE data. Second, we estimate the marginal *WTP* (*mWTP*) for the healthiness and carbon footprint of lasagne and compare these across the elicitation mechanisms using the Poe, Giraud and Loomis' convolution approach (2005).[15]

We analyse the SPVA data using generalised least-square regression models with correction for heteroscedasticity following Akaichi, Nayga and Nalley (2017) (Model 1a) (Equation 1)[16]:

[12] Priors estimated from the pilot study were used here. A small pilot for the SPVA with 10 subjects as well as 10 individual DCE-based interviews were conducted in January 2015. Subjects participating to the pilot were consumers and monetary incentives were not used in the pilot sessions.
[13] Choice tasks have been randomized to control for a potential order effect.
[14] To avoid food waste, the binding choice set was randomly drawn before the experiment. This draw was video recorded and the video was presented to subjects at the end of the experiment.
[15] Summary statistics of variable used to detect possible differences in sample composition across treatment groups are available in Appendix 4.
[16] As suggested by an anonymous referee, we test normality, homoscedasticity and correlation of the errors. While we reject normality and homoscedasticity, we do not detect correlation. Results are reported in Appendix 2.

$$BID_HG_{i,q} = \alpha + \beta_{HEA_A} HEA_A_{i,q} + \beta_{HEA_G} HEA_G_{i,q} + \beta_{CF_A} CF_A_{i,q} + \beta_{CF_G} CF_G_{i,q} + \gamma X + \varepsilon_{i,q} \quad (1)$$

The dependant variable ($BID_HG_{i,q}$) indicates the difference between individual i's bids for each lasagne and individual i's bid for the lasagne that is labelled as red in both healthiness and carbon footprint.[17] The dependent variable ranges from £-4 to £5. Our subjects are informed that the price of lasagne in the market ranges from £1 to £5. Our modelling approach does not control for observed bids being censored by the market prices, there are several reasons why we don't do this. First, the observed bids are not right censored at £5 because lasagne available in the market are not perfect substitutes of our experimental lasagne (see Harrison et al., 2004). Lasagne alternatives that are sold in Scottish retail stores do not show carbon footprint labels. Therefore, our subjects may be willing to bid more than £5 for a low carbon footprint lasagne. Second, observed bids are not left censored at £1 because our subjects can bid less than £1 to get a lasagne at a lower than market price. We find no evidence of bids being concentrated around the signalled prices, which suggests that the market price information did not affect subjects' bidding behaviour (see Figure 1 Appendix 3).[18] We estimate models that account for censoring to check the robustness of our results. Details are provided in Appendix 3.

The coefficient $\beta_{HEA_A,}$ denotes the average *mWTP* for an improvement in healthiness from red to amber, coefficient $\beta_{HEA_G,}$ denotes the average *mWTP* for an improvement in healthiness from red to green, coefficient $\beta_{CF_A,}$ denotes the average *mWTP* for an improvement in carbon footprint from red to amber, variable coefficient β_{CF_G} denotes the

[17] Following Gracia, Loureiro and Nayga (2011), we have $q = \{1,\ldots,8\}$.
[18] We expected any potential effect to be rather negligible because pre-cooked frozen lasagne is a very popular ready meal in the UK and therefore we did not tell our subjects something that they were not aware of. For example, the 43% of our sample eat this product at least 1 or 2 times a month, and 54% has eaten at least a lasagne in the last 3-4 weeks.

average *mWTP* for an improvement in carbon footprint from red to green. The term α is the slope and $\varepsilon_{i,t}$ is our error term.

A vector of interaction terms (*X*) between main variables and a set of socio-economic variable is incorporated in the modelling to control for differences in sample composition. These socio-economic variables were detected estimating a binary logit model where the dependent variable is whether each subject *i* belongs to the SPVA treatment group or not.[19] These socio-economic variables are: i) the variable *FREQ*, which indicates the number of times that a subject consumes lasagnes per month and is interacted with preferences for healthiness (*HEA_A*, *HEA_G*) and carbon footprint (*CF_A* and *CF_G*); ii) the variable *TASTE_HEA*, which indicates subjects' perceived tastiness of lasagne that are red compared to those that are green in healthiness (on a scale from 1 and 7). This variable is interacted with preferences for healthiness (*HEA_A*, *HEA_G*); iii) the variable *TASTE_CF*, which indicates the subjects' perceived tastiness of lasagne that are red compared to those that are green in in carbon footprint (on a scale from 1 and 7). This variable is interacted with preferences for carbon footprint (*CF_A* and *CF_G)*; iv) the variable *UNEMPL*, which indicates unemployment and is interacted with *HEA_A*, *HEA_G*, *CF_A* and *CF_G*.[20]

A random-parameters logit model in willingness-to-pay space is estimated using data collected from the DCE (Model 2a) (Train, 2009). The utility function of Model 2a is specified as follows (Equation 2):

$$V_{i,j,k} =$$

$$\beta_j + \beta_{HEA_A} HEA_A_{i,j,k} + \beta_{HEA_G} HEA_G_{i,j,k} + \beta_{CF_A} CF_A_{i,j,k} + \beta_{CF_G} CF_G_{i,j,k} +$$

[19] Results on differences in sample composition are available in Appendix 4. We acknowledge that there may be differences between subsamples that we were not able to control for.

[20] Another potential confounding could be the appearance of lasagne, even though lasagne look very similar. We asked subjects who participated to the extra 4 sessions in September 2017 whether they believed the lasagnes to look the same or not. Only one third of the subsample believe that the lasagne look different. We also run an additional model controlling for this effect and results suggest that appearance does not affect preferences (see Appendix 7).

$$\beta_{PR,i} PR_{i,j,k} + \gamma X \qquad (2)$$

In Equation 2, β_j is a coefficient indicating subjects' preferences for the opt-out alternative with respect to the lasagne alternatives.[21] The coefficients $\beta_{HEA_A,i}$ and $\beta_{HEA_G,i}$ indicates subjects' *mWTP* for lasagnes that are amber and green in healthiness (*HEA_A* and *HEA_G*, respectively) compared to lasagnes that are red in healthiness (*HEA_R*). The coefficients $\beta_{CF_A,i}$ and $\beta_{CF_G,i}$ indicate subjects' *mWTP* for lasagnes that are amber and green in carbon footprint (*CF_A* and *CF_G*, respectively) compared to lasagnes that are red in carbon footprint (*CF_R*). The coefficients β_{HEA_A} ,β_{HEA_G}, β_{CF_A} and β_{CF_G} are all assumed to be normally distributed with means and standard deviations to be estimated. The coefficient β_{pr} indicates subjects' preferences for the price of lasagnes (*PR*) and is modelled as a random parameter following a log-normal distribution with mean and standard deviation to be estimated. We also incorporate a set of interaction terms (*X*) between lasagne attributes and socio-economic variables that control for differences in sample composition across treatments. This is equivalent to set of interaction terms (*X*) for the SPVA data in Equation 1, except for the interaction terms incorporating the variable *UNEMPL*. In Equation 2, the variable *UNEMPL* is directly interacted with the lasagne price, *PR*.[22]

Assuming that the vector of coefficients θ associated to our random parameter is distributed according to the distribution $f(\theta)$, the probability *P* of choosing an alternative *j* is described in Equation 3:

$$P_{i,j,t} = \int \frac{e^{V_{i,j,t}}}{\sum_{j=1}^{J} e^{V_{i,j,t}}} f(\theta) d\theta \qquad (3)$$

[21] Lasagne alternatives are normalized to 0 to guarantee identification.
[22] We acknowledge that we do not estimate interaction between the main effect of the lasagne attributes and this may potentially bias the calculation of *WTPs* for lasagne types (Meas et al 2014).

The probabilities described in Equation 3 are numerically approximated by using methods of maximum simulated likelihood relying on 1,000 Halton draws (Train, 2009).

We acknowledge that other possible modelling approaches are available, for example the estimation of random parameter models in preference space. Therefore, to check the robustness of our results, we also estimated: i) a random parameter multinomial logit model where the price coefficient is assumed to be log-normally distributed and all other attributes are fixed; and ii) a random parameter multinomial logit model with where coefficients related to healthiness and carbon footprint of lasagnes are assumed to be normally distributed. These models are reported in Appendix 5. In the paper, we focus on the WTP space estimation because it allows us to directly estimate *mWTP* for the lasagne attributes (Scarpa, Thiene and Train, 2008). This makes SPVA and DCE data analyses strictly comparable.

We test whether SPVA and DCE are isomorphic by comparing $mWTP_L$ for healthiness and carbon footprint estimated from the DCE and SPVA treatments where $L = \{HEA_A, HEA_G, CF_A, CF_G\}$. We use the Poe, Giraud and Loomis' (2005) convolution approach. We use parametric bootstrapping techniques (i.e., Krinsky and Robb, 1986), to generate 1,000 bootstrapped values for each $mWTP_L$'s distribution and calculate 1,000,000 differences between the two bootstrapped distributions. We test the following hypotheses:

Hypothesis 1

H_0: $mWTP_{L,SPVA} = mWTP_{L,DCE}$

H_1: $mWTP_{L,SPVA} \neq mWTP_{L,DCE,}$

Elicitation mechanisms are isomorphic if and only if we fail to reject the null hypothesis (H_0).

4.2. Results

Table 2 reports the results of the generalised least-square regression models with correction for heteroscedasticity of SPVA bids (Model 1a).[23] We find that the *mWTP* for an improvement in healthiness from red to amber is £0.738 (β_{HEA_A}, $p<0.01$), and from red to green is £1.188 (β_{HEA_G}, $p<0.01$). The *mWTP* for an improvement carbon footprint from red to amber is £0.693 (β_{CF_A}, $p<0.05$), and from red to green is £1.126 (β_{CF_G}, $p<0.01$).[24,25]

[23] Tables providing main statistics of all variables used in our models are presented in Appendix 6.
[24] In Appendix 6, we also provide: i) results from censored and uncensored random effects models are presented in Appendix 6; ii) results from all models estimated without controlling for differences in treatment groups; iii) results from Hausman Test (1978) and Breush Pagan Lagrangian Multiplier Test.
[25] Appendix 7 provides results on the "appearance" and "instruction" effects.

Table 2. Generalised least-square regression models with correction for heteroscedasticity for SPVA Data[a]

	Model 1a		Model 1a
Dep.Var.:	***BID_HG***	**Dep.Var.:**	***BID_HG***
Variable	**Coefficient**	**Variable**	**Coefficient**
α	-0.271***	$\beta_{HEA_A_TASTE}$	-0.136***
	(0.0975)		(0.0221)
B_{HEA_A}	0.738***	$\beta_{HEA_G_TASTE}$	-0.303***
	(0.115)		(0.0221)
B_{HEA_G}	1.188***	$\beta_{CF_A_TASTE}$	0.0885
	(0.115)		(0.0583)
B_{CF_A}	0.693***	$\beta_{CF_G_TASTE}$	0.0569
	(0.114)		(0.0583)
β_{CF_G}	1.126***	$\beta_{HEA_A_UNEMPL}$	-0.0486
	(0.114)		(0.121)
$\beta_{HEA_A_FREQ}$	0.0147	$\beta_{HEA_G_UNEMPL}$	0.0127
	(0.0978)		(0.121)
$\beta_{HEA_G_FREQ}$	-0.0572	$\beta_{CF_A_UNEMPL}$	-0.218*
	(0.0978)		(0.121)
$\beta_{CF_A_FREQ}$	-0.109	$\beta_{CF_G_UNEMPL}$	-0.450***
	(0.0998)		(0.121)
$\beta_{CF_G_FREQ}$	-0.260***	*Log-likelihood*	-635.456
	(0.0998)	*Observations*	504
		Subjects	63

Note: ***p<0.01; **p<0.05; *p<0.10
[a] Standard Errors in parentheses

Table 3 reports the results of the Random Parameter Logit model estimated in WTP space using DCE choice-data (Model 2a).[26] We find that the *mWTP* for an improvement in healthiness from red to amber is £0.509 ($mWTP_{HEA_A}$) and *mWTP* for an improvement in healthiness from red to green is £1.022 ($mWTP_{HEA_G}$). Regarding carbon footprint, the *mWTP* for an improvement from red to green is £1.560 ($mWTP_{CF_G}$), while subjects are not willing to pay a premium for an improvement from red to amber in the carbon footprint dimension ($mWTP_{CF_A}$ = £0.05, not statistically different from £0).

[26] Appendix 8 provides: i) results from random parameters logit model estimated in preference space; ii) results from estimation of a multinomial logit estimated in WTP and preference space.

Table 3. Random Parameter Logit model estimated in WTP space using DCE choice-data[a,b,c]

	Model 2a		Model 2a
Dep.Var.:	***CHOICE_HG***	**Dep.Var.:**	***CHOICE_HG***
Variable	**Coefficient**	**Variable**	**Coefficient**
β_{OPT_OUT}	0.780***	$\beta_{HEA_A,\ MEAN,\ FREQ}$	-0.425***
	(0.223)		(0.148)
$\beta_{PRICE,\ MEAN}$	-0.247	$\beta_{HEA_G,\ MEAN,\ FREQ}$	0.342***
	(0.0928)		(0.0966)
$\beta_{HEA_A,\ MEAN}$	0.509***	$\beta_{CF_A,\ MEAN,\ FREQ}$	-0.125
	(0.124)		(0.0991)
$\beta_{HEA_G,\ MEAN}$	1.022***	$\beta_{CF_A,\ MEAN,\ FREQ}$	0.148
	(0.126)		(0.117)
$\beta_{CF_A,\ MEAN}$	0.0563	$\beta_{HEA_A,\ MEAN,\ TASTE}$	0.116***
	(0.149)		(0.0414)
$\beta_{CF_A,\ MEAN}$	1.560***	$\beta_{HEA_G,\ MEAN,\ TASTE}$	-0.0615**
	(0.170)		(0.0293)
$B_{PRICE,\ SD}$	1.924***	$\beta_{CF_A,\ MEAN,\ TASTE}$	0.171**
	(0.426)		(0.0669)
$\beta_{HEA_A,\ SD}$	1.171***	$\beta_{CF_G,\ MEAN,\ TASTE}$	-0.138**
	(0.162)		(0.0652)
$\beta_{HEA_G,\ SD}$	0.705***	$B_{PRICE,\ MEAN,\ UNEMPL}$	-0.136**
	(0.093)		(0.0543)
$\beta_{CF_A,\ SD}$	0.288***		
	(0.097)	*Log Likelihood*	-380.567
$\beta_{CF_G,\ SD}$	1.033***	Observations	1,755
	(0.117)	Subjects	65

Note: ***p<0.01; **p<0.05; *p<0.10
[a] Standard Errors in parentheses
[b] Allowing correlation
[c] In Model 2a, *PRICE* is a random parameter assumed to be lognormally distributed, and *HEA_A*, *HAE_G*, *CF_A* and *CF_B* are random parameters assumed to be normally distributed

Mean and confidence intervals of the $mWTP_L$'s empirical distributions and the results of Poe, Giraud and Loomis' test are reported in Table 4. We find that SPVAs and DCEs are not isomorphic. In fact, we find that the premium subjects are willing to pay for lasagnes that are green in healthiness, amber in healthiness and green in carbon footprint are statistically different in the SPVA and DCE (p<0.01, p<0.05 and p<0.01, respectively). These results are robust across model specifications as shown in Appendix 9.

Table 4. Comparisons of Marginal Willingness to Pay (in £) across Treatments[a]

	Model 1a	Model 2a	P-value
$mWTP_{HEA_A}$	0.738***	1.013***	0.160
	(0.446, 0.929)	(0.447; 1.442)	
$mWTP_{HEA_G}$	1.184***	2.041***	0.001
	(0.907; 1.379)	(1.343; 2.466)	
$mWTP_{CF_A}$	0.696***	0.102	0.033
	(0.392;0.888)	(-0.624; 0.600)	
$mWTP_{CF_G}$	1.129***	3.124***	0.000
	(0.852; 1.315)	(2.135; 3.684)	

Note: ***p<0.01; **p<0.05; *p<0.10
[a] Confidence interval (5%; 95%) is reported in parentheses

5. Testing value elicitation problems using IV procedures

5.1. Testable hypotheses and model specifications

Using data from our IV tasks, we test whether the lack of isomorphism in HG preferences is a value-elicitation problem. In a SPVA, the response strategy that maximizes subjects' payoff is to bid an amount of money equal to the resale value associated with the token *t*. Therefore, SPVA is demand revealing if the bids ($BID_IV_{i,t}$) are equal to resale values ($IV_{i,t}$). In the DCE, subjects face a choice *k* between two tokens and doing nothing. Each token has a pay-off, which is the difference between the price and the resale value of the token. A DCE is demand revealing if observed choices ($CH_IV_{i,k}$) are equal to choices that maximizes subjects' payoff, hereafter expected choices ($EXP_CH_IV_{i,k}$).

We investigate demand revelation at both the aggregate and individual level. At aggregate level, in the SPVA treatment, we calculate the percentage of bids in which the subjects maximize their payoff (i.e., demand revealing bids). In the DCE tasks, we measure the percentage of payoff maximising choices (i.e., demand revealing choices). The degree of difficulty of the two elicitation mechanisms may differ. We allow for this by defining the degree of demand revelation in the SPVA treatment group in two ways: i) the SPVA is *demand revealing with no error margin*, if the submitted bid for a token is equal to the resale

value associated to that token; and ii) the SPVA is *demand revealing with ± £0.50 margin error*, if the submitted bid for a token is £0.50 greater or smaller to the resale value associated to that token. This approach follows that of Shogren *et al.* (2001) and is consistent with our IVs which were randomly drawn from a uniform distribution of [£1.00, £5.00] in £0.50 increments.

To analyse demand revelation at the individual level, we estimate two econometric models. Model 3a is estimated using SPVA data, and Model 4a DCE data. Model 3a (i.e., SPVA) is estimated using a two-way random-effects procedure. A tobit model is not appropriate because data are not censored. In fact, subjects were not forced to submit positive bids in the SPVA treatment (see experimental instructions in the supplementary appendix). Model 3a takes the same form of Equation 4.

$$BID_IV_{i,t} = \alpha + \beta IV_{i,t} + \varphi_t + \mu_i + \varepsilon_{i,t} \quad (4)$$

The dependant variable ($BID_IV_{i,t}$) is individual i's bids for each token $t = \{1,..,T\}$ and ranges from £0 to £7. Our variable $IV_{i,t}$ denotes subject i's induced value for token t, α is the intercept, the term μ_i represents subject-specific characteristics, the terms φ_t are introduced as dummies in the models and represent token-specific effects such as learning or fatigue trends in bidding behaviour[27] and $\varepsilon_{i,t}$ is our error term with mean zero and variance σ_ε. These models are estimated under the assumptions that μ_i and φ_t are drawn from a bivariate normal distribution.[28]

The SPVA is demand revealing if we fail to reject the null joint hypothesis that the estimated parameter α is equal to zero, the coefficient associated to the IV variable (β) is

[27] Only T-1 terms are estimated as token 1 is used as baseline to guarantee identification.
[28] The modelling approach used for data collected in the SPVA treatment replicates that used by Shogren *et al.* (2001) and Parkurst, Shogren and Dickinson (2004).

equal to 1, and the vector of coefficients denoting token-specific effects (φ_t) are equal to zero in Model 3a (as in Shogren et *al.*, 2001). In fact, if these conditions are met Equation 4 reduces to $BID_IV_{i,t} = IV_{i,t}$. Therefore, the following hypothesis is tested here using standard post estimation procedures (i.e., Wald Test).

Hypothesis 2

H_0: $\alpha = 0$, $\beta = 1$ and $\varphi_t = 0$

H_1: $\alpha \neq 0$, $\beta \neq 1$ and $\varphi_t \neq 0$

Model 4a (DCE) is modelled using Random Utility Models (RUMs). The utility that subject *i* attaches to each alternative *j* in each choice set *k* is decomposed into two parts, $V_{i,j,k}$, the part of the utility observed by the researcher, and $\varepsilon_{i,j,t}$, which cannot be observed by the researcher, so that, $U_{i,j,k} = V_{i,j,k} + \varepsilon_{i,j,k}$. In our application, we assume that $\varepsilon_{i,j,k}$ are iid extreme value and therefore we estimate a Multinomial Logit Model. In both models, the utility function is specified as follows (Equation 6):

$$V_{i,j,k} = \beta_j + \beta_{IV} IV_{i,j,k} + \beta_{PR} PR_{i,j,k} \quad (5)$$

In Equation 5, β_j is a coefficient indicating subjects' preferences for the token alternatives with respect to the opt-out.[29] The coefficients β_{IV} and β_{PR} inform us on preferences for the selling ($IV_{i,j,k}$) and buying price ($PR_{i,j,k}$), respectively. Variables $IV_{i,j,k}$ and $PR_{i,j,k}$ indicate induced-value and buying prices of each alternative j presented in our choice sets.

The estimation of multinomial logit is problematic, if a large majority of subjects

[29] Only J-1 terms are estimated because the alternative specific constant associated to the opt-out alternative is dropped to guarantee identification.

always choose the dominant option (in our case the payoff maximising option).[30] However, we do not find evidence of this: 66.5 percent of choices are for the payoff maximising option and only the 38.5 percent of participants always select the payoff maximising option. Nevertheless, we explore alternative modelling approaches that mitigate this issue: i) error component logit model; ii) scale heterogeneity multinomial model (Fiebig et al., 2009); and iii) generalized multinomial logit model (Fiebig et al., 2009) or scaled mixed logit model (Greene and Hensher, 2010).

We assume that a choice-based elicitation mechanism is demand revealing if we fail to reject the null hypothesis that the estimated coefficients β_{IV} and the negative of β_{PR} are equal to each other. Therefore, the following hypotheses are tested:

Hypothesis 3

H_0: $\beta_{IV} = -\beta_{PR}$

H_1: $\beta_{IV} \neq -\beta_{PR}$

5.2. Results

At the aggregate level, we find that 31.92% of bids in the SPVA are equal to the IVs, and, hence, strictly demand revealing. When we allow for ±£0.50 error margin, this percentage rises to 52.57%. We find most subjects underbid (60.01%) and this is in line with previous findings by Noussair, Robin, and Ruffieux (2004), Shogren et *al.* (2001) and Lusk and Rousu (2006). In our DCE treatment, the percentage of demand revealing (payoff maximizing) choices is 65.58%, which is similar to the results reported in Luchini and Watson (2014).

[30] We thanks an anonymous referee for pointing this out.

Table 6 reports the results of the two-way random-effects models that test demand revelation in the SPVA treatment (Model 3a). We find that the coefficient α and most of the terms φ_t are not statistically different from 0, while the coefficient β is equal to 0.859 (p<0.01). In a Wald Test we reject the null hypothesis that $\alpha = 0$, $\beta = 1$ and φ_t=0 (*Hypotheses 2*) (p<0.05), and, hence, we conclude that SPVA is not demand revealing.[31] We find that $BID_IV_{i,t} = \alpha + \beta IV_{i,t} + \varphi_t + \mu_i + \varepsilon_{i,t}$ (Equation 4) does not reduce to $BID_IV_{i,t} = IV_{i,t}$.

[31] Results from One-way and Two-way Fixed Effects models using data collected in the SPVA treatment are reported in Appendix 10 along with those from the Hausman Test (1978) and the Breusch and Pagan Lagrangian Multiplier tests. These testing procedures show that Random-Effects Models are superior to Fixed-Effects and Ordinary Least Square models.

Table 6. Random-Effects Models for the SPVA Treatment[a,b]

	Model 3a
Dep .Var.:	***BID_IV***
Variable	**Coefficient**
α	0.0849
	(0.112)
β	0.859***
	(0.0227)
φ_{TOK2}	-0.0485
	(0.125)
φ_{TOK3}	-0.140
	(0.125)
φ_{TOK4}	-0.390***
	(0.125)
φ_{TOK5}	-0.271**
	(0.125)
φ_{TOK6}	-0.340***
	(0.125)
φ_{TOK7}	-0.163
	(0.125)
φ_{TOK8}	0.0427
	(0.125)
φ_{TOK9}	-0.0392
	(0.125)
σ_u	0.979
$\sigma_{\varepsilon,}$	0.758
ρ	0.664
H_0:α=0 & β=1 & φ_{TOKt}=0	38.86***
R-squared	0.478
Observations	567
Subjects	63

Note: ***p < 0.01, **p < 0.05, *p < 0.1

[a] Robust standard errors in parentheses - clustering

[b] Wald test

Table 7 reports the results from all the choice models estimated to test demand revelation in DCE treatment: i) multinomial logit models (Model 4a); ii) error component logit model (Model 4b); ii) scale heterogeneity multinomial model (Model 4c); and iii) generalized multinomial logit model or scaled mixed logit model (Model 4d). Using a Wald Test, we fail to reject the null hypothesis $\beta_{IV} = -\beta_{PR}$ (*Hypothesis 3*) in all estimated model specifications and we conclude that the DCE is demand revealing.

At the individual level we find that the DCE is the most demand revealing preference elicitation technique. We conclude that DCE is more demand revealing than SPVA, and therefore, the lack of isomorphism found is a value-elicitation problem.

Table 7. Models for DCE Treatment[a,b,c]

	Model 4a	**Model 4b**	**Model 4c**	**Model 4d**
Dependent Variable	***CH_IV***	***CH_IV***	***CH_IV***	***CH_IV***
Variable	**Coefficient**	**Coefficient**	**Coefficient**	**Coefficient**
β_{IV}	3.489***	3.905***	3.760***	3.905***
	(0.286)	(0.266)	(0.729)	(0.747)
B_{PR}	-3.489***	-3.871***	-3.731***	-3.871***
	(0.242)	(0.245)	(0.669)	(0.687)
$\beta_{OPT\text{-}OUT,\ MEAN}$	-0.251	-0.276	-0.198	-0.276
	(-0.294)	(0.284)	(0.283)	(0.292)
$\beta_{OPT\text{-}OUT,\ ST.DEV.}$	-	1.002***	-	1.002***
	(-)	(0.209)	(-)	(0.323)
τ	-	-	0.0587*	-0.000984
	(-)	(-)	(0.0328)	(0.00432)
H_0:$B_{IV} = -\beta_{PR}$	0.000	0.200	0.130	0.180
Log likelihood	-295.651	-277.356	-283.356	-277.350
Observations	1,755	1,755	1,755	1,755
Number of id	65	65	65	65

Note: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$
[a] Standard errors in parentheses
[b] Model 4a is multinomial logit model
Model 4b is an error component model
Model 4b is scale heterogeneity multinomial logit
Model 4c is a generalized multinomial logit model
[c] Wald test

6. Testing value-formation by combining IV and HG procedures

6.1. Testable hypotheses and model specifications

In this section, we use a novel method to test whether lack of isomorphism between *mWTP* estimates for lasagne attributes elicited using a SPVA add DCE is entirely a value-elicitation problem. We test if HG preferences are equal across the SPVA and DCE when only subjects who provide demand revealing bids and choices in the IV tasks are included in the analysis. Demand revealing bids and choices are those maximizing subjects' payoffs in the IV SPVA and DCE tasks, respectively. If only those subjects, who submit demand revealing bids and choices, are included in the statistical analysis, then any differences we find in *mWTP*s are not due to a value-elicitation problem and instead indicate a value formation problem.

We test this hypothesis using the same two-step procedure, modelling approaches and estimation procedures implemented to test isomorphism. First, we analyse the SPVA data using generalised least-square regression models with correction for heteroscedasticity (Model 5a) and the DCE data using a random parameters logit model estimated in WTP space (Model 6a). Then, we compare $mWTP_{L,DR}$ - the *mWTP* for the lasagne attributes estimated only for 'demand revealing' respondents - across the SPVA and DCE. We use the Poe, Giraud and Loomis' (2005) convolution approach to test the following hypotheses:

Hypothesis 4

H_0: $WTP_{L,SPVA,DR} = WTP_{L,DCE,DR}$

H_1: $WTP_{L,\,SPVA,DR} \neq WTP_{L,DCE,DR}$

Value-formation is ruled out if and only if we fail to reject the null hypothesis (H_0).

6.2. Results

In the SPVA, 14 (out of 63) subjects submitted only demand revealing bids in the IV tasks. In the DCE, 25 (out of 65) subjects always made demand revealing choices in the IV tasks.[32] Table 8 presents results for the SPVA (Model 5a), while Table 9 presents the results from the DCE (Model 6a).[33,34,35] Table 10 presents a comparison of *mWTPs* that are estimated for lasagne attributes using SPVA and DCE data and suggests that $mWTP_{L,DR}$ for some lasagne attributes differ across treatment groups (*Hypotheses* 4). We find differences in *mWTP* for a lasagne that is amber in healthiness ($p<0.01$), amber in carbon footprint ($p<0.05$) and green in carbon footprint ($p<0.01$) between SPVA and DCE treatment groups.

These results suggest that lack of isomorphism detected in our experiment is not solely due to a value elicitation problem in the SPVA. In this analysis all respondents were able to make demand revealing bid in the IV SPVA or choices in IV DCE. This implies that our finding of differences in *mWTP* for lasagne attributes across the SPVA and DCE may also be due to value-formation problem. These results are robust across model specifications as shown in Appendix 15.[36] However, we acknowledge that caution is needed when interpreting these results because sample sizes in our treatment are now very small. Further research with a larger sample size is needed.

[32] Results on distributions of subjects making demand revealing bids or choices are presented in Appendix 11. Interestingly, there are no subjects submitting 7 or 8 demand revealing bids or choices (out of 9), indicating that our sample of subjects in both treatments make either demand revealing or non-demand revealing bids/choices consistently.

[33] Appendix 12 shows that normality and homoscedasticity are not supported (while autocorrelation is not an issue (Table 6).

[34] Appendix 13 provides: i) results from estimation of random effects censored and uncensored models; iii) results from the estimation of all models without control for differences in subsample composition; iii) results from the estimation of models assuming the ±0.5 error margin.

[35] Appendix 14 provides: provides: i) results from random parameters logit model estimated in preference space; ii) results from estimation of a multinomial logit estimated in WTP and preference space.

[36] Appendix 15 also presents comparisons of *mWTPs* across treatment groups when we take into account a ±£0.50 error margin in the SPVA IV task.

Table 8. Generalised least-square regression models with correction for heteroscedasticity for SPVA Data[a]

	Model 5a		**Model 5a**
Dep.Var.:	***BID_HG***	**Dep.Var.:**	***BID_HG***
Variable	**Coefficient**	**Variable**	**Coefficient**
α	-0.051	$\beta_{HEA_A_TASTE}$	-0.270***
	(0.184)		(0.0644)
β_{HEA_A}	-0.857**	$\beta_{HEA_G_TASTE}$	-0.522***
	(0.432)		(0.0644)
β_{HEA_G}	1.281***	$\beta_{CF_A_TASTE}$	-0.168
	(0.432)		(0.126)
B_{CF_A}	0.351	$\beta_{CF_G_TASTE}$	-0.244*
	(0.298)		(0.126)
β_{CF_A}	0.473	$\beta_{HEA_A_UNEMPL}$	1.457***
	(0.298)		(0.309)
$\beta_{HEA_A_FREQ}$	0.389	$\beta_{HEA_G_UNEMPL}$	0.263
	(0.252)		(0.309)
$\beta_{HEA_G_FREQ}$	-0.542**	$\beta_{CF_A_UNEMPL}$	0.0204
	(0.252)		(0.237)
$\beta_{CF_A_FREQ}$	-0.237	$\beta_{CF_G_UNEMPL}$	-0.222
	(0.254)		(0.237)
$\beta_{CF_G_FREQ}$	-0.174	*Log-likelihood*	-96.711
	(0.254)	*Observations*	112
		Subjects	14

Note: ***p<0.01; **p<0.05; *p<0.10

[a] Standard Errors in parentheses

Table 9. Random parameter multinomial logit model in WTP space for DCE Data[a,b,c]

	Model 6a		Model 6a
Dep.Var.:	***CHOICE_HG***	**Dep.Var.:**	***CHOICE_HG***
Variable	**Coefficient**	**Variable**	**Coefficient**
β_{OPT_OUT}	3.968***	$\beta_{HEA_A,\ MEAN,\ FREQ}$	-0.300
	(0.721)		(0.293)
$\beta_{PRICE,\ MEAN}$	-0.877	$\beta_{HEA_G,\ MEAN,\ FREQ}$	1.118***
	(0.952)		(0.312)
$\beta_{HEA_A,\ MEAN}$	0.429*	$\beta_{CF_A,\ MEAN,\ FREQ}$	0.212
	(0.256)		(0.273)
$\beta_{HEA_G,\ MEAN}$	0.663***	$\beta_{CF_A,\ MEAN,\ FREQ}$	-0.187
	(0.247)		(0.294)
$\beta_{CF_A,\ MEAN}$	0.877**	$\beta_{HEA_A,\ MEAN,\ TASTE}$	-0.209
	(0.351)		(0.151)
$\beta_{CF_A,\ MEAN}$	2.130***	$\beta_{HEA_G,\ MEAN,\ TASTE}$	-0.0608
	(0.476)		(0.128)
$B_{PRICE,\ SD}$	2.460***	$\beta_{CF_A,\ MEAN,\ TASTE}$	-0.0224
	(0.713)		(0.127)
$\beta_{HEA_A,\ SD}$	-1.575***	$\beta_{CF_A,\ MEAN,\ TASTE}$	0.104
	(0.244)		(0.193)
$\beta_{HEA_G,\ SD}$	2.334***	$B_{PRICE,\ MEAN,\ UNEMPL}$	0.542***
	(0.441)		(0.205)
$\beta_{CF_A,\ SD}$	-0.0302		
	(0.0958)	*Log Likelihood*	-138.879
$\beta_{CF_A,\ SD}$	1.875***	Observations	675
	(0.367)	Subjects	25

Note: ***p<0.01; **p<0.05; *p<0.10

[a] Standard Errors in parentheses

[b] Allowing correlation

[c] In Model 2a, *PRICE* is a random parameter assumed to be lognormally distributed, and *HEA_A*, *HAE_G*, *CF_A* and *CF_B* are random parameters assumed to be normally distributed

Table 10. Comparisons of Marginal Willingness to Pay (in £) across Treatments[a]

	Model 6a	Model 6a	P-value
$mWTP_{HEA_A,\ DR}$	-.870**	0.872	0.005
	(-2.031; -0.169)	(-0.522; 1.760)	
$mWTP_{HEA_G,\ DR}$	1.294***	1.289***	0.495
	(0.169; 2.011)	(0.139; 2.082)	
$mWTP_{CF_A,\ DR}$	0.340	1.742***	0.030
	(-0.465; 0.817)	(-0.057; 2.292)	
$mWTP_{CF_G,DR}$	0.467	4.294***	0.000
	(-0.396; 0.950)	(1.732; 5.924)	

Note: ***p<0.01; **p<0.05; *p<0.10
[a] Confidence interval (5%; 95%) is reported in parentheses

7. Conclusion

Previous experiments eliciting HG preferences for food products have shown that auction- and choice-based approaches are not isomorphic, meaning that the use of different elicitation mechanisms leads to different WTP estimates for the same good. An open question remains whether this is because these elicitation mechanisms differ in their degree of demand revelation and/or because HG preferences are formed differently across elicitation mechanisms.

This paper is the first study to answer this question using an innovative experimental design that combines IV and HG procedures for a SPVA and DCE. We elicit IV preferences for a fictitious good (token) and HG preferences for real food product (beef-based lasagne) that varies in healthiness and carbon footprint. Our study design allows us to test isomorphism and determine whether this phenomenon is a value-elicitation problem. In addition, it allows also testing for a value-formation problem while controlling for value elicitation differences.

We find that HG preferences elicited using SPVAs and DCEs are not isomorphic. After controlling for potential differences in sample composition, our results suggest that HG preference patterns varies depending on the elicitation mechanism. Our investigation of IV

preferences shows that the DCE is the most demand revealing preference-elicitation procedure. Taken together, our results imply that that lack of isomorphism in our empirical application is caused by a value-elicitation problem in the SPVA, but this result does not rule out value-formation problems in the HG task as well.

We investigate value formation-issues by comparing the HG preferences and *mWTP*s of subjects submitting only demand revealing bids or making only demand revealing choices in the IV tasks of our SPVA and DCE treatments, respectively. We find that estimated *mWTP*s varies across mechanisms, and hence, we conclude that the lack of isomorphism in our HG preferences is also due to a value formation problem.

Future wok could explore whether value-elicitation issues occur because SPVA is not demand revealing for some non-expected utility maximizer subjects, as suggested by Horowitz (2006), or because the experimental designs induce subjects to fail to identify the optimal bidding strategy (i.e. random mistakes).

Our experiment is also informative about consumers' preferences for health- and environmental-related food traits. To the best of our knowledge, this is the first attempt to investigate consumers' trade-offs between healthy and environmental sustainable versions of food products. As our IV analysis suggests that DCE provides more accurate preferences, we focus on results from the estimation results of our random-parameter logit model in WTP space. These results suggest that consumers prefer green to red lasagnes in both dimensions, while they only prefer amber to red lasagne when it comes to healthiness. Consumers value substantial reductions (i.e., green) in carbon footprint more than reduction in healthiness, while the opposite is true for moderate reductions (i.e., amber). Results also suggest that SPVA underestimates the premiums subjects are truly willing to pay for healthier and more environmental sustainable lasagnes.

References

Akaichi, F., Nayga, R. and Nalley, L.L. (2017). Are there trade-offs in valuation with respect to greenhouse gas emissions, origin and food miles attributes? *European Review of Agricultural Economics* 44(1):3-31.

Becker, G.M., DeGroot, M.H. and Marschak J. (1964). Measuring Utility by a Single-Response Sequential Method. *Behavioural Science* 9:226-32.

Balcombe, K., Fraser, I. and Falco S.D. (2010). Traffic lights and food choice: A choice experiment examining the relationship between nutritional food labels and price. *Food Policy* 35(3):211-220.

Caputo, V., Nayga, R. and 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.

Carbon Trust 2012. *Carbon Footprinting Guide*. March 2012.

Carlsson-Kanyama, A. and González, A.D. (2009). Potential contributions of food consumption patterns to climate change. *The American Journal of Clinical Nutrition* 89(5):1704-1709.

Carson, R. and Groves, T. (2007). Incentive and information properties of preference questions. *Environmental and Resource Economics* 37:181–210.

ChoiceMetrics, (2011). Ngene 1.1 User manual and reference guide. http://www.choice-metrics.com/download.html [accessed 11.23.2015]

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.

Coppinger, V.M., Smith, V. and Titus, L. (1980). Incentives and Behavior in English, Dutch and Sealed-Bid Auctions. *Economic Inquiry* 18(1):1-22.

Cox, J., Roberson, B. and Smith, V. (1982). Theory and behavior of single object auctions. In V. L. Smith, ed. *Research in experimental economics* (vol. 2). Greenwich CT: JAI Press.

Cummings, R.G., Elliott, S., Harrison, G.W. and Murphy, J. (1997). Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy* 105(3):609-621.

Fiebig, D.G., Keane, M.P., Louviere, J. and Wasi, N. (2009). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science* 29(3): 393 – 421.

Food Standards Agency (2013). *Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets*. 19th June.

Gracia, A., Loureiro, M. and Nayga, R.M.Jr. (2011). Are valuations from non-hypothetical choice experiments different from experimental auctions? *American Journal of Agricultural Economics* 93(5):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.

Green, W.H. and Hensher, D.A. (2010). Does Scale Heterogeneity across Individual Matter? An Empirical Assessment of Alternative Logit Models. *Transportation* 37: 413-428.

Harrison, G.W. (2006). Making choice studies incentive compatible. In B.J. Kanninen, ed. *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*. Dordrecht, The Netherlands: Springer.

Harrison, G.W. and List, J.A. 2004. Field Experiment. *Journal of Economic Literature* 42(4):1009-1055.

Harrison, G.W., Harstad R.M. and Rutsröm, E.E. (2004). Experimental Methods and Elicitation of Values. *Experimental Economics* 7(2): 123–140.

Hausman, J.A. (1978). Specification Tests in Econometrics. *Econometrica* 46(6): 1251-1271.

Horowitz, J.K. (2006). The Becker-DeGroot-Marschak Mechanism is not Always Incentive Compatible, even for Non-Random Goods. *Economic Letters* 93:6-11.

Kagel, J., Harstad, R. and Levin, D. (1987). Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica* 55:1275–1304.

Kagel, J., and Levin, D. (1993). Independent private value auctions: Bidder behaviour in first-, second- and third-price auctions with varying numbers of bidders. *The Economic Journal* 103: 868–879.

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

Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics* 68(4): 715–719.

Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy* 74:132–157.

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

Lusk, J.L., and Rousu, M. (2006). Market price endogeneity and accuracy of value elicitation mechanisms. In J.A. List, ed. *Using Experimental Methods in Environmental and Resource Economics*. Northampton US: Edward Elgar.

Lusk, J.L. and Shogren, J.F. (2007). *Experimental Auctions. Methods and Applications in Economic and Marketing Research.* Cambridge (UK): Cambridge University Press.

Lusk, J.L. and Schroeder, T. (2006). Auction bids and shopping choices. *Advances in Economics Analysis and Policy* 6(1):1-37.

Macdiarmid, J.I., Douglas, F. and Campbell, J. (2016). Eating like there's no tomorrow: public awareness of the environmental impact of food and reluctance to eat less meat as part of a sustainable diet. *Appetite* 96: 487-493.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. in P. Zarembka, ed. *Frontiers in Econometrics*. New York US: Academic Press.

Meas, T., Hu, W., Batte, M. T., Woods, T. A. and Ernst, S. (2014). Substitutes or complements? Consumer preference for local and organic food attributes. *American Journal of Agricultural Economics* 97(4): 1044-1071.

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.

Parkhurst, G.M., Shogren, J. F. and Dickinson, D.L. (2004). Negative Values in Vickrey Auctions. *American Journal of Agricultural Economics* 86(1): 222–235.

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(2): 353-365.

Rutsröm, E.E. (1998). Home-grown values and incentive compatible auction design. *International Journal of Game Theory*, 27:427-441.

Scarpa, R., Thiene, M. and 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.

Shogren, J., Margolis, M., Koo, C. and List, J. (2001). A random nth price auction. *Journal of Economic Behavior and Organization* 46(4):409–421.

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

Train, K. (2009). *Discrete Choice Methods with Simulation*, 2nd. ed. Cambridge, UK: University Press.

Vickrey, W. (1961). Counterspeculation, Auctions and Competitive Sealed Tenders. *Journal of Finance* 16:8-37.

Vossler, C.A. and McKee, M. (2006). Induced-Value Tests of Contingent Valuation Elicitation Mechanisms. *Environmental & Resource Economics* 35(2): 137-168.