European Review of Agricultural Economics Vol **40 (2)** (2013) pp. 379–404
doi:10.1093/erae/jbs030
Advance Access Publication 20 August 2012

Preference reversals in Contingent and Inferred valuation methods

Spiros Stachtiaris[†], **Andreas C. Drichoutis**[‡]* **and Stathis Klonaris**[†]

[†]*Agricultural University of Athens, Greece;* [‡]*University of Ioannina, Greece*

Received August 2011; final version accepted April 2012

Review coordinated by Paolo Sckokai

Abstract

We examine inconsistencies in preference orderings using the Contingent valuation (CV) and the Inferred valuation (IV) methods. We find that in the context of a food market we do not observe strong inconsistencies. Weak inconsistencies are observed for the IV method, indicating that IV is slightly more susceptible to inconsistent preference orderings than the CV method. We also find that the IV method generates higher valuations than CV in the case of consumers with high commitment costs (that is, low familiarity with the product) but successfully mitigates social desirability bias in the case of low commitment costs and high normative motivations.

Keywords: willingness to pay, Contingent valuation, Inferred valuation, preference reversals

JEL classification: C93, D12

1. Introduction

Eliciting people's valuation for non-market goods has been central in the economics literature. The Contingent valuation (CV) method is by far the most popular method of valuing the benefits of a new good in monetary terms by estimating individuals' maximum willingness to pay (WTP). The CV method is not based on actual observation of a market value. Rather, it uses valuations elicited from hypothetical surveys and thus suffers from a number of documented biases. This method was principally developed in environmental and transport economics, but over the last decades has made considerable headway in the valuation of food products, such as organic products (Buzby *et al.*, 1998), pesticide-free fruits and vegetables (Boccaletti and

*Corresponding author: Department of Economics, University of Ioannina, University Campus, 45110 Ioannina, Greece. E-mail: adrihout@cc.uoi.gr

Nardella, 2000), origin labelled wines (Skuras and Vakrou, 2002), novel food products (Brummett and Nayga, 2007), new quality products (Corsi, 2007), organic wine (Brugarolas *et al.*, 2009a), traditional horticultural varieties (Brugarolas *et al.*, 2009b) and certified foods (Hsu *et al.*, 2009).

Because of hypothetical responses in the CV method, there is a widespread belief among economists that there is little incentive for consumers to truthfully reveal their preferences and in turn researchers tend to distrust people's ability to answer CV questions accurately (List and Gallet, 2001; Little and Berrens, 2004; Murphy *et al.*, 2005). As a result, it is often impossible to determine whether the choices/answers made in a hypothetical survey are consistent with how the individual would behave if actually given the opportunity to do so. Therefore, the lack of actual economic commitments can lead to hypothetical bias in which CV overestimates the true economic value of the good. To this extent, Alfnes and Rickertsen (2007) showed how to extrapolate experimental auction results using stated choice surveys, whereas Alfnes, Yue and Jensen (2010) tested a method they called 'real talk' to mitigate hypothetical bias based on cognitive dissonance theory. However, since most, if not all, CV studies are conducted in hypothetical contexts,[1] results from this and other studies that look into refining the empirical methodology are important in their own right.

Recently, Lusk and Norwood (2009a) developed a new method that instead of asking people what *they* are willing to pay, asks subjects what *they think another (average) person would pay*.[2] This form of indirect questioning was first proposed by Fisher (1993) who found that indirect questioning affected responses in questions with normative context but not in questions with personal motivation. Respondents, in the presence of an interviewer, may report socially desirable preferences and thus misrepresent their 'true' preferences, in order to either please the interviewer or to be consistent with social norms (Crowne and Marlowe, 1960; Fisher, 1993; Leggett *et al.*, 2003; Plant, Devine and Brazy, 2003; List *et al.*, 2004). The respondent wishes to provide the answer that is most 'socially acceptable' rather than speak his/her true feelings. Thus, the presence of an interviewer may lead to social desirability bias, whereby respondents provide responses that they think will please the interviewer or be consistent with society norms (Leggett *et al.*, 2003). Moreover, Lusk, Pruitt and Norwood (2006) found that subjects exhibit stronger moralistic and pro-social behaviour when they know they are being scrutinised.

Lusk and Norwood (2009a) coined the term 'Inferred valuation' (IV) to describe this type of indirect questioning in valuation studies. The aim of the IV method is not only to alleviate social desirability, but also to moderate hypothetical bias. The implicit assumption is that social desirability bias is part of hypothetical bias, and therefore by eliminating or reducing social desirability

1 Particularly true for environmental valuation studies where a real market with salient payments is difficult to establish.

2 A similar concept was introduced in Cummings and Harrison (1992).

bias via the IV method we can also alleviate (part of) hypothetical bias. The extent to which we can achieve that will depend on the extent to which social desirability bias and hypothetical bias overlap.

The contribution of this study to the (agricultural) economics literature is on the methodological front. The ultimate contribution to the research community is to develop methods that would accurately measure values without having to develop a market for a good. The evaluation of the IV method in terms of its prediction accuracy and consistency to rational choice theory, using the CV method as a benchmark, helps us perform this service. This paper therefore seeks to critically re-assess the potential benefits of the IV method over the CV method as a means to elicit WTP in food markets. Our approach for re-assessing these two methodologies is discussed in Section 2. Section 3 presents our experimental design for testing our hypothesis, whereas Section 4 presents our hypotheses and results. We conclude in the last section.

2. Theoretical approach and objectives of the study

A core element of economic theory posits that individuals choose the best action according to stable and unchanging preferences (Becker, 1976: 5). For psychologists (and behavioural economists), this basic assumption of economics is far away from the truth; preferences may be formed at the point a question is asked and can be highly sensitive to the manner in which the question is asked. Thus, labile, malleable and/or context-sensitive preferences are the scientific obvious facts and are not considered a bias as in standard rational choice theory.

In order to uncover preferences, CV involves asking the respondent a (sequence of) question(s); answers, if truthful, are direct expressions of preferences, which possibly include normative or moral considerations. On the other hand, the simple twist in the wording of the valuation question in Lusk and Norwood (2009b) generated (inferred) valuations that are possibly free of normative or moral aspects. In fact, the IVs elicited in Lusk and Norwood (2009b) were close to real valuations (when compared with an experiment) and lower than hypothetical valuations (where social desirability is prevalent). With the IV method, individuals are asked to predict how other people would behave and thus infer *other people's preferences* that are ideally free from normative or moral considerations.

The fact that the IV method is capable of alleviating social desirability bias was also confirmed in Lusk and Norwood (2009a). Their emphasis was on exploring the role of normative motivations and the gap between the laboratory and the field. The authors showed that goods with normative dimensions are more prone to social desirability bias and thus the IV method is more effective in bridging the gap between the laboratory and field valuations. Thus, the authors identified conditions under which the IV method might be more successful. Moreover, the role of familiarity with the product and commitment costs were central in our analysis as well, since Lusk and Norwood (2009a)

found that people understated their preferences with the IV method for relatively familiar goods with normative attributes.[3]

So how can we test whether and when elicited preferences are consistent with rational choice theory, if at all? The way we adopt in this study is to investigate whether or not CV and IV refute the basic assumption of the rational choice theory: preferences are consistent and stable. The often-cited strand of the literature that deals with non-consistent preference orderings is the preference-reversal literature (see Seidl, 2002 for a review). Broadly defined, any systematic change in preference orderings between normatively equivalent conditions can be called a preference reversal (Slovic and Lichtenstein, 1983). The preference-reversal literature took off with the study of Slovic and Lichtenstein (1968) and the help of economists (Grether and Plott, 1979) that demonstrated the robustness of the effect. The phenomenon is an empirical regularity such that a pricing task for lotteries reveals opposite preferences from a choice task made out of the lotteries. Preference-reversal studies typically ask subjects to choose between pairs of lotteries (e.g. lottery A vs. lottery B) and then ask subjects to price lotteries in an auction-type mechanism. Ideally, if a person chooses A over B in a choice task, s/he should also price lottery A higher than lottery B in an auction task, otherwise we observe a preference reversal.

More recently, List (2002) (as well as Alevy, List and Adamowicz, 2011) demonstrated a different type of preference reversals: those that occur between joint and isolated evaluation modes (as opposed to different elicitation methods, e.g. the pricing and choice task mentioned above). Preference reversals of the 'more-is-less' type, a term made popular by List (2002), are based on the fact that the value of a consumer good may change when it is evaluated alone compared with when it is evaluated jointly with another product.[4] Valuations of two related products are compared, where one of which clearly dominates the other; the dominance is varied by having goods valued jointly and in isolation between subjects. Preference reversals of the 'more-is-less' type occur when the dominated good is valued more highly in the isolated mode. List (2002) showed that preferences in the sports card market follow a 'more-is-less' pattern. Whereas in a joint evaluation mode, a superior bundle of sports cards ('More' product) is consistently valued more highly than an inferior bundle ('Less' product), in an isolated mode, the inferior bundle is valued more than the superior bundle of cards. He also showed that the effect is significantly attenuated for experienced subjects.

Hsee (1996) (as well as Hsee *et al.*, 1999) proposed the evaluability hypothesis as an explanation for preference reversals between valuation

3 Commitment cost is the 'option value' associated with gaining additional information about the product and the ability to postpone the purchase. We follow this line of argument later in our analysis by assuming that people's previous purchases of goods in a product category are an appropriate proxy of commitment costs (Lusk and Norwood, 2009a). When people buy a product more often, commitment costs are low.

4 This may be an indication on why retailers pay such close attention to product displays.

modes. He suggested that preference reversals between joint and isolated evaluations occur because one of the attributes involved in the options is hard to evaluate independently and another attribute is relatively easy to evaluate independently. When these attributes are presented jointly, evaluation is facilitated. In fact, Hsee (1996) showed that when *both* attributes are hard to evaluate or easy to evaluate, preference reversals disappear.

The consequences of preference reversals are significant since they refute the basic assumption of the rational choice theory that preferences are consistent and stable. In contrast, they back up a behavioural decision theory which states that preferences are constructed on the spot when asked to form a particular judgment or to make a specific decision (Slovic, 1995; Payne, Bettman and Schkade, 1999; Johnson, Steffel and Goldstein, 2005; Lichtenstein and Slovic, 2006). In this sense, values are not merely uncovered when elicited; they are partly constructed at that time. The constructive viewpoint does not necessarily mean that there is no 'true' value to be measured but that expressed preferences reflect *both* uncovered preferences and the particular heuristics used to construct the required response under the specific elicitation mechanism (Payne, Bettman and Schkade, 1999).

Therefore, the question that arises is whether predicting preferences with the IV method could result in more consistent and well-defined preference orderings when compared with standard preference elicitation methods such as CV. To answer this question, we move into the context of a food market instead of the sports cards market used in List (2002).[5] This is highly relevant and prompt since many studies seek to value new food products and a significant part of the valuation literature is filled with such studies. It is therefore worth knowing whether the uncovered preferences using elicitation methods such as CV or IV are well defined and consistent with the assumptions of economic theory. In addition, our study allows us to re-evaluate the effectiveness of the IV method in mitigating social desirability bias.

3. Experimental design

3.1. Significant choices of the experimental design

Before proceeding with the specifics of our experiment, we first discuss two issues that might be considered significant choices of the experimental design. The first choice has to do with the issue of choosing between a within-subjects or a between-subjects design. In a 'within-subjects' (WS) design, each individual is exposed to more than one treatment being tested, whereas in a 'between-subjects' (BS) design, each individual is exposed to only one

5 As one of the reviewers noted, in hypothetical markets, errors of valuation are not costly. With public goods, a mistaken valuation results in a big monetary outlay, whereas with market goods the error in valuation can be small. If the hypothetical context of our study precluded subjects from thinking carefully about their valuations, then the IV method may be more prone to error if the IV method is harder for subjects to provide a valuation than the CV method. Our methods do not allow us to refute this point but this may be a prime area for future research.

treatment. Both the WS and BS designs have advantages and disadvantages. We chose to perform a BS design with respect to the CV and IV treatments. This has several advantages when compared with a WS design. For example, in WS designs, subjects may feel more compelled to differentiate their answers between scenarios (Frederick and Fischhoff, 1998). Further on this point, Charness, Gneezy and Kuhn (2012) describe an example of a WTP elicitation scenario in a WS design which varies the order of the valuation scenarios between individuals. However, varying the order of the scenarios may result in additional biases since values for the second scenario will always be biased by exposure to the first scenario. Simply stated, subjects always have a reference or comparison point when responding to the second question.

In addition, Charness, Gneezy and Kuhn (2012) argue that random assignment can be a powerful tool that researchers need to trust for producing usable results in the BS analysis, and this is the case with our study as well. An additional caveat for applying a WS design in our study is that in one of the treatments it is necessary to elicit four different valuations for two different products. If we were to proceed with a WS design for the CV and IV treatments, then we would need to significantly increase (double) the number of elicited valuations from the same subject. This might have resulted in fatigue effects, early termination of the interview or even 'yeah-saying' bias.

The second major choice in our experiment is that we chose a payment card as our valuation elicitation format. In the CV literature, it is generally reported that using different elicitation formats results in different WTP values. The open-ended format was criticised by the National Oceanographic and Atmospheric Administration (NOAA; Arrow *et al.*, 1993) as providing 'erratic and biased' responses. The NOAA panel suggested the dichotomous choice (DC) format which became the favoured approach for several years. Other alternatives that were not considered by the NOAA panel were later developed including the multiple bounded dichotomous choice (MBDC) and the payment card formats. However, several studies document that iterative question formats produce anomalies in respondent behaviour such as anchoring problems where the final bid at the end of the iterations[6] was found to be significantly correlated with the starting bid (Thayler, 1981; Boyle, Bishop and Welsh, 1985; Boyle, Johnson and McCollum, 1997; Green *et al.*, 1998). Herriges and Shogren (1996) have shown that a double-bounded question format (MBDC) exacerbates anchoring because of the fact that responses to the second value are influenced by the magnitude of the first one. There is evidence to suggest that open-ended and dichotomous choice surveys do not demonstrate convergent validity due to 'yeah-saying' bias (Kanninen, 1995) which is the tension of some respondents to answer affirmative to any value presented to them (Holmes and Kramer, 1995). Carson and Groves (2007),

6 An iterative question format starts by asking respondents, 'would you pay X €?' for a specified good or policy. If respondents answer yes, then the bid is increased until they say no. On the other hand, if the initial response was no, the bid is decreased until the respondents say yes. The starting bids, extent of bid iterations and number of iterations (repetitions) vary depending on the case.

also argue against the binary discrete choice format when the introduction of a new private good is concerned.

The payment card format exhibits more desirable properties than DC or MBDC (Reaves, Kramer and Holmes, 1999), less 'yeah-saying' at high bid amounts (Zhongmin *et al.*, 2006) and results in more conservative estimates (Blaine *et al.*, 2005). It is not a surprise therefore that in health economics it is by far the most common format for CV studies (Smith and Sach, 2010). Donaldson, Thomas and Torgerson (1997) argued in favour of the PC format for its resemblance to real-world behaviour, where individuals 'shop around', observe different values for a good and choose the one that suits them most. Thus, due to its resemblance to every-day behaviour, cognitive demand is potentially mitigated and the validity of the instrument is increased.

For all the above reasons, the payment card format has been widely used in food (Brummett and Nayga, 2007; Hsu *et al.*, 2009; Aizaki *et al.*, 2011; Hu *et al.*, 2011) and resource economics studies (Lienhoop and MacMillan, 2007; Brouwer, van Beukering and Sultanian, 2008; Solomon and Johnson, 2009; Simpson and Hanna, 2010; Ressurreição *et al.*, 2012). Following this line of research, we used a payment card format which consisted of 16 price intervals.[7] The prices were selected from pre-tests so as to cover a wide range of prices for the products we used. The payment card intervals were constructed using an exponential response scale to avoid range and centring bias that is prevalent in classical uniform payment cards (Rowe, Schulze and Breffle, 1996). Drichoutis, Lazaridis and Nayga (2009) describe the procedure of constructing an exponential card in detail.

Overall, the field experiment we designed was a BS experiment (subjects are exposed to either CV or IV scenario but not both) and it is similar in concept to List's (2002) and Alevy, List and Adamowicz's (2011) experiments. However, we do not require actual payment as in the original studies but we rather elicit hypothetical valuations. To the extent that hypothetical bias equally affects elicited valuations under different evaluation modes, results should remain unaffected. Therefore, several of the procedures for studying the implications of preference reversals across joint and isolated valuation modes were similar to the above cited studies. Moreover, we alter List's (2002) and Alevy, List and Adamowicz's (2011) studies by replacing the sport cards market with the food market. More than one food product (olive oil and apples) is used for each valuation method to check for the robustness of our results. Data were collected in supermarkets from consumers while shopping. For half of the respondents, valuations were elicited with the CV method and for the other half with the IV method. All valuation products were exhibited in photo stimuli.[8] Subjects were asked to report their maximum WTP for the good in the photo which was also described orally.

7 An example of the exact wording of the valuation questions are exhibited in Appendix D in Supplementary data at *ERAE* online.

8 The photo stimuli are available upon request from the authors.

3.2. Design issues

The field experiment was carried out in supermarkets located in Athens. The experimenter approached each participant and invited him/her to participate voluntarily in an interview. We approached 882 persons and got 593 agreements to participate in the survey and 289 refusals. Five persons terminated the interview early, thus were dropped from the analysis resulting in 588 valid responses. Thus, the response rate was 67.2 per cent and the co-operation rate was 66.2 per cent. If the respondent accepted the invitation, then she/he was randomly allocated to one of the two evaluation modes (joint or isolated) and to one of the two elicitation methods (CV or IV). If she/he was allocated to the isolated mode, she/he was then further allocated to the 'More' or 'Less' product. This 3 × 2 design is exhibited in Table 1.

In the joint mode, subjects evaluated the products 'Less' and 'More' simultaneously, whereas in the isolated mode subjects evaluated either the 'Less' or the 'More' product but not both (i.e. we had three treatments: (1) Less and More-joint, (2) More-isolated and (3) Less-isolated).

In each treatment, subjects were asked to evaluate two food products (olive oil and apples). Order was altered between subjects. The specific products used are exhibited in Table 2. Under the 'More' product, the superior quality products and the inferior counterpart were tied together and presented as a single product, thus we refer to the 'More' product as a single product from now on. Standard socio-demographic data were also collected.

Our full factorial design is a 3 × 2 × 2 design.[9] In all, it took 12 subjects to complete the full factorial design one time. An example is given in Table A1 (Appendix A in supplementary data at *ERAE* online). As exhibited, 12 subjects are required to participate in six treatments for two quality categories (BIO and PDO).[10]

To sum up, in the 'Isolated' evaluation modes, subjects report their valuation either for a quality food product (LI for Less-Isolated) or for a quality food product *tied with* a smaller quantity of a conventional product (MI for More-Isolated) (see Table 2). In the 'Joint' evaluation modes, subjects report their valuation for both the quality product (LJ for Less-Joint) as well as the quality product *tied together with* a conventional product (MJ for More-Joint). This design ran for two quality categories (BIO and PDO) and two elicitation methods (CV and IV).

We should note that, in aggregate, the superior food quality product *tied with* the lower quality product have a greater market value than the superior

9 Three evaluation modes (Less and More-joint vs. More-isolated vs. Less-isolated) × two elicitation methods (Contingent valuation vs. Inferred valuation) × two quality categories (BIO vs. PDO).

10 While one might argue that social desirability bias may not be important for small food purchases such as olive oil and apples, Lusk and Norwood (2009a) demonstrated that this can be an issue even for products with potentially small purchases. In their study, they used an 'environmentally friendly' dishwashing liquid, a 90 per cent lean organic ground beef product and an organic regionally grown whole wheat flour.

Table 1. Field experiments – experimental design

		Elicitation method	
Evaluation modes	Products (less or more)	Contingent valuation	Inferred valuation
Isolated	Less Isolated (LI)	LI-CV	LI-IV
	More Isolated (MI)	MI-CV	MI-IV
Joint	Less Joint (LJ)	LJ-CV	LJ-IV
	More Joint (MJ)	MJ-CV	MJ-IV

Table 2. Products by evaluation mode

	Treatments by evaluation modes		
	Less Isolated (LI)	More Isolated (MI)	Less Joint (LJ) and More Joint (MJ)
Olive oil			
Product 1	BIO olive oil (750 ml)		BIO olive oil (750 ml)
Product 2		BIO olive oil (750 ml) + conventional olive oil (250 ml)	BIO olive oil (750 ml) + conventional olive oil (250 ml)
Product 3	PDO olive oil (750 ml)		PDO olive oil (750 ml)
Product 4		PDO olive oil (750 ml) + conventional olive oil (250 ml)	PDO olive oil (750 ml) + conventional olive oil (250 ml)
Apples			
Product 5	BIO apples (1 kg)		BIO apples (1 kg)
Product 6		BIO apples (1 kg) + conventional apples (250 g)	BIO apples (1 kg) + conventional apples (250 g)
Product 7	PDO apples (1 kg)		PDO apples (1 kg)
Product 8		PDO apples (1 kg) + conventional apples (250 g)	PDO apples (1 kg) + conventional apples (250 g)

food quality product itself. In the 'Joint' evaluation mode, subjects evaluate the exact same products as in treatments LI and MI but this time side by side.

No subject participated in more than one treatment. In addition, subjects evaluated the products using either CV or IV method. Third, each subject reported his/her valuation for one quality category, either a PDO product or a BIO product but not both. However, each subject reported valuations for two food products, i.e. olive oil and apples. Lastly, order of appearance of valuation questions (and products) was altered between subjects.

3.3. The survey

WTP was elicited through two payment cards, one for each product, i.e. olive oil and apples (see Appendix B in supplementary data at *ERAE* online), in an actual marketplace just before subjects enter a supermarket. Interviews took place at various locations throughout the city, at stores of the three of the biggest food retailers in the country. The interviews were conducted by a single proctor (one of the authors) from Monday to Saturday, during morning and afternoon hours.[11] Table 3 depicts socio-demographic information from the sample.

List (2002) showed that preference reversals of the more-is-less type are significantly attenuated for experienced subjects. To distinguish between experienced and inexperienced people in our study, we asked subjects to self-rate their knowledge about either BIO or PDO products by asking them to indicate whether they agree with the statement 'I know about these products' on a 1–5 scale anchored by completely disagree and completely agree. Subjects that indicated to agree or completely agree with the above statement were categorised as familiar with the products and thus experienced.

Lusk and Norwood (2009a) showed that when it comes to the IV method, normative motivations (i.e. social concerns) and commitment costs (i.e. familiarity with the product) can be important determinants for IV to successfully narrow the laboratory–field gap. To isolate normative motivations for BIO and PDO products in our study, we asked subjects to indicate whether they agree with the statement 'I should be looking to purchase $<<$BIO or PDO$>>$ products' on a 1–5 scale anchored by completely disagree and completely agree. Subjects that indicated to agree or completely agree with the statement were categorised as having strong normative motivations (Lusk and Norwood, 2009a).

To proxy commitment costs, we asked respondents to indicate how often they tend to buy BIO or PDO olive oil and apple products on a Likert scale ranging from 1 to 5 (never, rarely, sometimes, often, always). Subjects that indicated buying often or always were classified as having low commitment costs.

With the IV method, subjects are asked to predict WTP for the 'average consumer'. Therefore, if the person being asked thinks she/he is less price-sensitive than the 'average' shopper, this may result in a lower WTP in the IV than in the CV. For this reason, price sensitivity was also recorded by

11 Supermarkets are closed on Sundays throughout the country.

Table 3. Variable description

Variables	Variable description	Mean (S.D.)
$Income_1^a$	Dummy, household's economic position is bad or very bad = 1	0.049
$Income_2$	Dummy, household's economic position is below average = 1	0.066
$Income_3$	Dummy, household's economic position is average = 1	0.505
$Income_4$	Dummy, household's economic position is above average = 1	0.197
$Income_5$	Dummy, household's economic position is good = 1	0.143
$Income_6$	Dummy, household's economic position is very good = 1	0.039
$Educ_1^a$	Dummy, education level is up to high school = 1	0.059
$Educ_2$	Dummy, education level is high-school graduate = 1	0.354
$Educ_3$	Dummy, education level is university graduate = 1	0.471
$Educ_4$	Dummy, education level is postgraduate = 1	0.115
Age	Subject's age	45.094 (12.440)
Child	Dummy, subject has underage children in household = 1	0.415
HSize	Household size	2.901 (1.381)
Gender	Dummy, male = 1	0.349
Exper	Dummy, 1 = experienced subject	0.423
Norm	Dummy, 1 = strong normative motivation	0.325
PrSens	Dummy, 1 = subject is price sensitive	0.811
Commit	Dummy, 1 = subject has low commitment cost in buying olive oil	0.207
	Dummy, 1 = subject has low commitment cost in buying apples	0.252

Note: Standard deviation is provided for non-dummy variables.
[a]Variables were omitted from the econometric models.

having subjects answer a 5-point Likert question regarding how often they take price under consideration while grocery shopping. Answers were anchored by never and always and subjects that answered often or always were classified as price-sensitive.

4. Hypotheses and results

4.1. Descriptive statistics of the WTP responses

Before proceeding with testing our hypotheses, insights can be gained by looking at some descriptive statistics of the WTP responses. Figures 1 and 2 (3 and 4) show the distribution of responses over the 16-cell payment cards for olive oil (apples) under the CV and IV method, respectively.

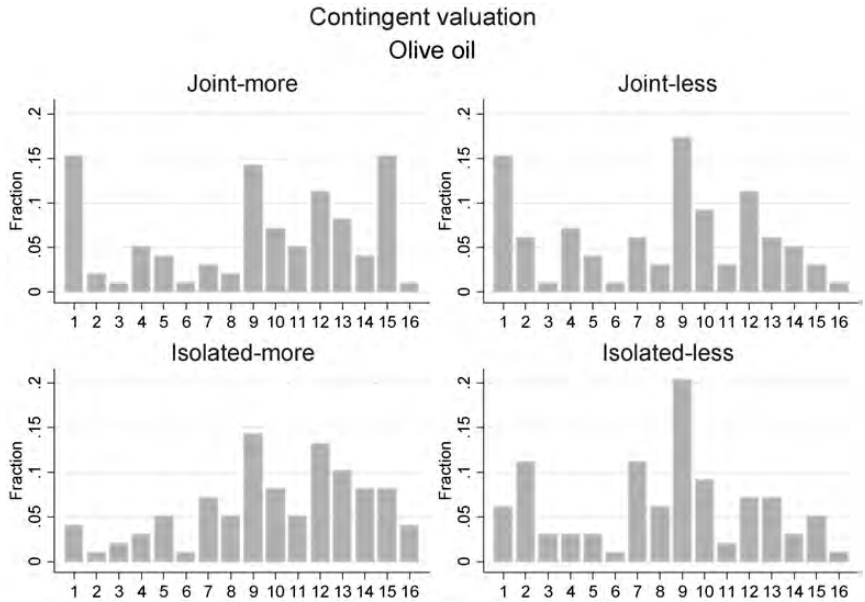

Fig. 1. Distribution of responses by payment card cells for olive oil with CV.

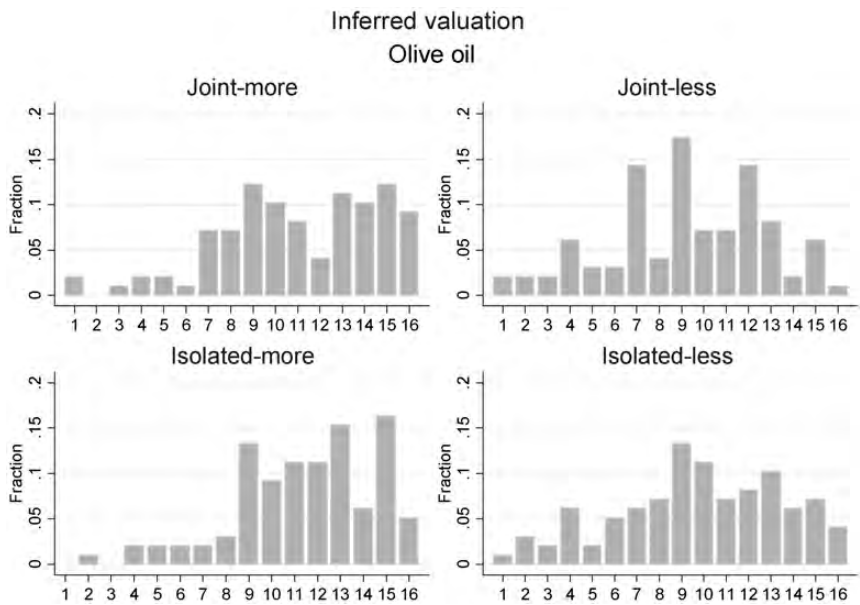

Fig. 2. Distribution of responses by payment card cells for olive oil with IV.

If one closely observes the figures, it is apparent that for the 'Joint' mode, the distribution of responses for the 'More' product is shifted more to the right, when compared with the distribution of responses for the 'Less' product. This

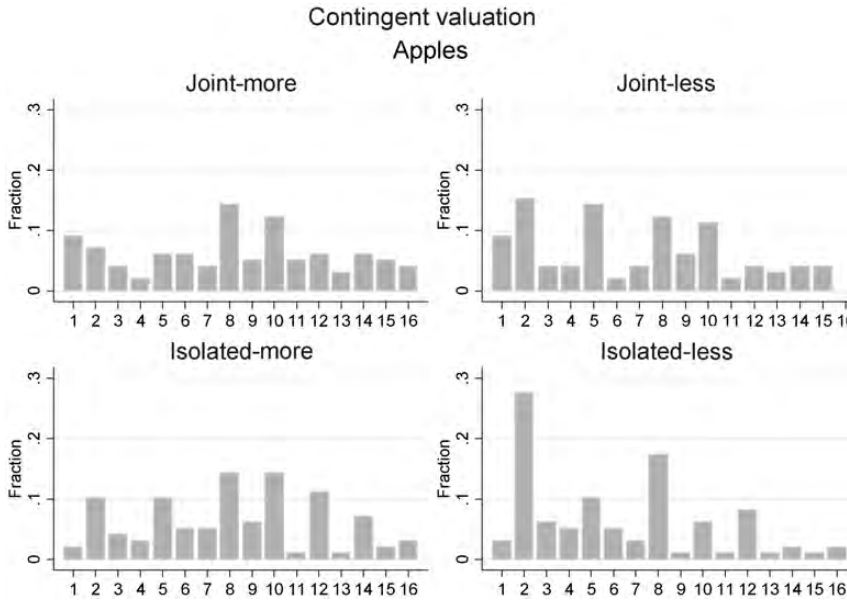

Fig. 3. Distribution of responses by payment card cells for apples with CV.

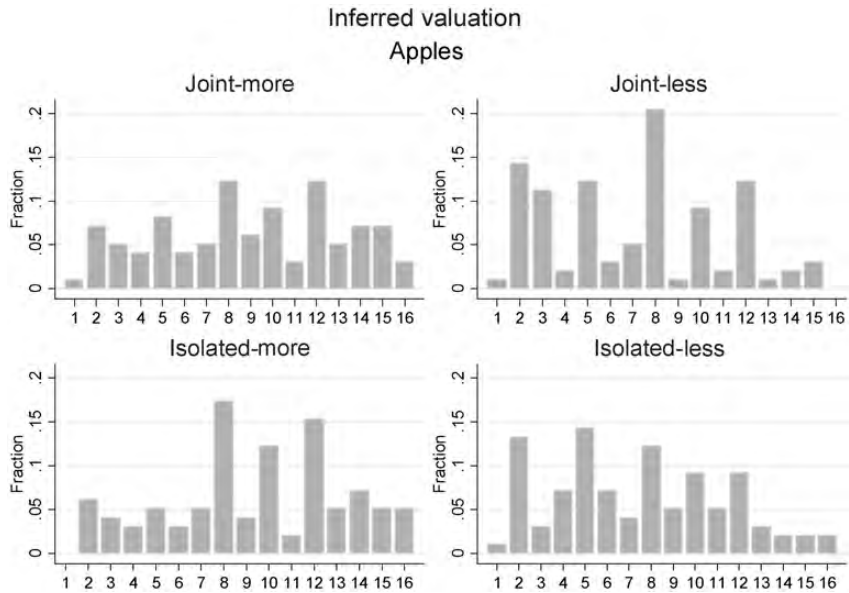

Fig. 4. Distribution of responses by payment card cells for apples with IV.

indicates that WTP for the 'More' product is greater than that for the 'Less' product when products are evaluated jointly. Surprisingly, there is a similar pattern for the 'Isolated' mode indicating the absence of a preference reversal

(implying consistent preference orderings). In addition, there is a clear shift of the distribution of responses to the right for the IV method when compared with the CV method; this is irrespective of the evaluation mode (Joint or Isolated). This indicates that valuations elicited with the IV method appear to be greater than valuations elicited with the CV method which refutes our basic assumption that IV mitigates social desirability bias.

Before moving to the conditional analysis, it is also important to investigate how the goods and subjects vary along key variables, i.e. experience of the subject, normative motivations and commitment costs. Overall results indicate that people felt more experienced with BIO products (mean = 3.43) than PDO products (mean = 2.91). p-Values from a Wilcoxon–Mann–Whitney (WMN) test ($p = 0.0$) indicate that there is a significant difference between degree of familiarity for the two product categories.

Results also indicate that people felt higher normative motivations for BIO products (mean = 2.81) than PDO products (mean = 2.55). p-Values from a WMN test ($p = 0.02$) indicate that the difference is statistically significant. With respect to commitment costs, subjects indicated lower commitment costs for PDO olive oil (mean = 2.11) than BIO olive oil (mean = 1.69) and the difference is statistically significant according to a WMN test ($p = 0.001$).[12] However, for apples, it is the exact opposite since subjects indicate lower commitment costs for BIO apples (mean = 2.45) than PDO apples (mean = 1.91) and the difference is statistically significant according to a WMN test ($p = 0.001$).

4.2. Hypotheses

To formally test our hypothesis for preference reversals, we adopt the definitions from Alevy, List and Adamowicz (2011):

Definition 1: A strong evaluation mode effect is observed when, in aggregate, preferences over the goods are: LI (Less, Isolated) $\succ$ MI (More, Isolated) and MJ (More, Joint) $\succ$ LJ (Less, Joint).

Definition 2: A weak evaluation mode effect is observed when, in aggregate, preferences over the bundles are: LI $\sim$ MI and MJ $\succ$ LJ.

To test the effect of IV on elicited valuations, we can directly test whether Inferred $<$ Contingent.

Table 4 summarises the test forms that we adopt to test for preference reversals across the CV and IV methods.[13] To test these hypotheses, we estimate an

12 Note that a higher value in the scale indicates lower commitment costs and vice versa.

13 Detailed derivations are available from the authors upon request. See the Appendix in Supplementary data at *ERAE* online for more details.

Table 4. Linear combinations of coefficients for hypothesis testing for preference reversals

Hypothesis tested	Test form		Elicitation method or valuation mode	
Preference reversals	MI $\prec$ LI and MJ $\succ$ LJ	Experienced[a]	Contingent	More + More × Exper < 0 and More + More × Joint + More × Exper + More × Joint × Exper > 0
			Inferred	More + More × Infer + More × Exper + More × Infer × Exper < 0 and More + More × Joint + More × Infer + More × Exper + More × Joint × Infer + More × Joint × Exper + More × Infer × Exper + More × Joint × Infer × Exper > 0
		Inexperienced[a]	Contingent	More < 0 and More + More × Joint > 0
			Inferred	More + More × Infer < 0 and More + More × Infer + More × Joint + More × Joint × Infer > 0

[a]The expressions involved in the last column concerns coefficients which take the name of their respective dummies.

interval regression model (to take into account the interval nature of the dependent variable) with robust clustered standard errors (to account for multiple responses by the same person in the 'Joint' modes). Table C1 (Appendix C in supplementary data at *ERAE* online) shows the various linear combination forms that we use to test whether Inferred $<$ Contingent.[14] The empirical specification follows closely Alevy, List and Adamowicz's (2011) specification:

$$\begin{aligned}\mathrm{WTP}_i = {} & a_1 + a_2\,\mathrm{More}_i + a_3\,\mathrm{Joint}_i + a_4\,\mathrm{Infer}_i + a_5\,\mathrm{Exper}_i + a_6\,\mathrm{BIO}_i \\ & + a_7\,\mathrm{Norm}_i + a_8\,\mathrm{Commit}_i + a_9\,\mathrm{PrSens}_i + a_{10}\,\mathrm{BIO}_i \times \mathrm{Exper}_i \\ & + a_{11}\,\mathrm{BIO}_i \times \mathrm{Norm}_i + a_{12}\,\mathrm{BIO}_i \times \mathrm{Commit}_i i + a_{13}\,\mathrm{BIO}_i \times \mathrm{PrSens} \\ & + a_{14}\,\mathrm{Infer}_i \times \mathrm{Norm}_i + a_{15}\,\mathrm{Infer}_i \times \mathrm{Commit}_i + a_{16}\,\mathrm{Infer}_i \times \mathrm{PrSens}_i \\ & + a_{17}\,\mathrm{Infer}_i \times \mathrm{Exper}_i + a_{18}\,\mathrm{More}_i \times \mathrm{Joint}_i + a_{19}\,\mathrm{More}_i \times \mathrm{Infer}_i \\ & + a_{20}\,\mathrm{More}_i \times \mathrm{Exper}_i + a_{21}\,\mathrm{Joint}_i \times \mathrm{Infer}_i + a_{22}\,\mathrm{Joint}_i \times \mathrm{Exper}_i \\ & + a_{23}\,\mathrm{More}_i \times \mathrm{Joint}_i \times \mathrm{Infer}_i + a_{24}\,\mathrm{More}_i \times \mathrm{Joint}_i \times \mathrm{Exper}_i \\ & + a_{25}\,\mathrm{More}_i \times \mathrm{Infer}_i \times \mathrm{Exper}_i + a_{26}\,\mathrm{Joint}_i \times \mathrm{Infer}_i \times \mathrm{Exper}_i \\ & + a_{27}\,\mathrm{More}_i \times \mathrm{Joint}_i \times \mathrm{Infer}_i \times \mathrm{Exper}_i \\ & + a_{28}\,\mathrm{OrderQuest}_i + b_1\,\mathrm{dem}_{1i} + b_2\,\mathrm{dem}_{2i} + \cdots + u_i. \end{aligned} \tag{1}$$

The dem_j variables are a series of demographic variables described in Table 3.[15] The More, Joint, Infer, Exper, BIO, Norm, Commit and PrSens variables are dummies indicating conditions consistent with the variable name, i.e. evaluation of the 'More' product (vs. the 'Less' product), evaluation in the 'Joint' mode (vs. the 'Isolated' mode), evaluation using the IV method (vs. CV), experienced consumers (vs. inexperienced consumers), evaluation of BIO (vs. PDO), subject has high normative motivations for the product (vs. low normative motivations), subject has low commitment costs with the product (vs. high commitment costs) and subject is price-sensitive (vs. no price sensitivity), respectively. The OrderQuest variable is a dummy controlling for the order of the valuation questions between food products (olive oil and apples).

4.3. Does CV and IV generate consistent preference orderings?

To answer this question, we test the hypotheses as described in Table 4. Notice that this test requires checking two hypotheses: a confirmation of inconsistent preference orderings requires that MI $\prec$ LI and MJ $\succ$ LJ, in

14 Detailed derivations are available from the authors upon request. See the Appendix in Supplementary data at *ERAE* online for more details.

15 Results from the raw interval regression output of equation (1) are available from the authors upon request.

Table 5. Hypothesis tests for preference reversals[a]

		Olive oil (p-value)*		Apples (p-value)*	
	Hypothesis tested	Experienced	Inexperienced	Experienced	Inexperienced
CV	H_0: WTP in the MI mode > WTP in the LI mode (isolated)	1.372 (0.998) MI ≻ LI	1.504 (0.999) MI ≻ LI	0.341 (0.990) MI ≻ LI	0.545 (0.999) MI ≻ LI
	H_0: WTP in the MJ mode < WTP in the LJ mode (joint)	0.571 (0.051) MJ ≻ LJ	1.081 (0.000) MJ ≻ LJ	0.080 (0.262) MJ ∼ LJ	0.467 (0.000) MJ ≻ LJ
IV	H_0: WTP in the MI mode > WTP in the LI mode (isolated)	0.787 (0.868) MI ∼ LI	1.400 (0.999) MI ≻ LI	−0.041 (0.432) MI ∼ LI	0.630 (0.999) MI ≻ LI
	H_0: WTP in the MJ mode < WTP in the LJ mode (joint)	1.572 (0.000) MJ ≻ LJ	1.112 (0.000) MJ ≻ LJ	0.314 (0.001) MJ ≻ LJ	0.319 (0.000) MJ ≻ LJ

Note: The implied preference relation is depicted just below the p-value.
[a]The numbers without parentheses are the sum of the corresponding coefficients provided in Table 4.
*p-Values in parentheses. Note that a high p-value for H_0 implies a low p-value for H_1. Therefore, a p-value >90 per cent or >95 per cent for H_0 would be equivalent to a rejection of H_1 at the 10 per cent or 5 per cent level, respectively.

aggregate. Table 5 shows the results of these tests. The interaction terms associated with the *Exper* dummy allows us to further disentangle the effect of market experience on preference reversals. For each food product (olive oil and apples) and elicitation method (CV and IV), we first test whether the respective linear combination of coefficients from Table 5 is ≥ 0 (H_0: Linear Comb ≥ 0). The alternative hypothesis (H_1: Linear Comb < 0) is consistent with MI $\prec$ LI. We then test whether the respective linear combination of coefficients is ≤ 0 (H_0: Linear Comb ≤ 0). The alternative hypothesis (H_1: Linear Comb > 0) is consistent with MJ $\succ$ LJ. Note that any p-value exhibited in the table implies ($1 - p$ value) for the alternative hypothesis. Implied preference relations are exhibited just below p-values in Table 5.

First notice that the majority of linear combinations of coefficients are evaluated as positive which implies that MI $>$ LI and MJ $>$ LJ. More specifically, most hypothesis involving H_0: WTP in the MI mode $>$ WTP in the LI mode cannot be rejected which suggests that average WTP in the MI mode is statistically significantly higher than average WTP in the LI mode. However, there are some exceptions: medium-sized p-values (or 10 per cent $< p$-value $<$ 9 per cent) indicate that we can reject neither H_0 nor H_1. In turn, this implies that MI $\sim$ LI. On the other hand, the hypothesis involving H_0: WTP in the MJ mode $<$ WTP in the LJ mode is rejected in all but one cases, implying that MJ $\succ$ LJ. Overall, our findings indicate that we never observe strong evaluation mode effects as was the case in List (2002) (as well as Alevy, List and Adamowicz, 2011). Note that direct comparisons should be done with caution since, in contrast to those studies, we did not offer actual payments for eliciting valuations.

However, we do observe weak evaluation mode effects (i.e. MI $\sim$ LI and MJ $\succ$ LJ). It is worth noting that weak preference reversals are observed only for the IV method and only for experienced subjects. There are two conclusions coming out of these results. The first one is that market experience does not play a significant role for the CV method. It appears that both experienced and inexperienced consumers were not likely to commit a preference reversal in aggregate. The second conclusion is related to the fact that we find weak preference reversals for experienced subjects in the IV method. This makes sense if we assume that experienced subjects are more likely to expect others to fall prey to social desirability bias and thus predict for others that, in aggregate, MI $\sim$ LI. On the other hand, inexperienced subjects may not expect other people to fall for social desirability bias and thus predict MI $\succ$ LI.

Overall, in contrast to List (2002) and Alevy, List and Adamowicz (2011), we cannot replicate strong evaluation mode effects. However, we do observe weak evaluation mode effects for the IV method. Therefore, the IV method is slightly more susceptible in generating inconsistent preference orderings than

the CV method. Market experience did not make much of a difference in our results, as in the original studies.[16]

4.4. Does IV generate lower valuations than CV?

The aim of the IV method, as originally used, was to mitigate social desirability bias that is encompassed in hypothetical bias. Lusk and Norwood (2009a, 2009b) found that for goods with high normative motivations, IV generated lower WTP valuations than people's own valuations. Therefore, for goods for which people have high normative motivations, we would expect average WTP from IV to be lower than average WTP from CV: Inferred $<$Contingent. On the other hand, when subjects have low normative motivations for the good, we would expect Inferred = Contingent (see also Figure 1 in Lusk and Norwood, 2009a). Table C1 (Appendix C in supplementary data at *ERAE* online) indicates linear combinations of coefficients that are required to test our hypothesis, by evaluation mode, subjects experience, normative motivations, commitment costs and price sensitivity. Results of these one-sided tests are displayed in Tables 6 and 7. A positive value for the linear combination of coefficients indicates that Inferred $>$ Contingent, whereas a negative value indicates the exact opposite.

Visual inspection of Tables 6 and 7 indicates that the linear combination of coefficients often results in a positive value, indicating that average WTP from the IV method is greater than average WTP from the CV method. However, Tables 6 and 7 also exhibit negative values, indicating that average WTP from IV is lower than that from CV. This is most often the case for inexperienced subjects and/or with high normative motivations and/or low commitment costs. p-Values can help us decide whether the observed differences are statistically significant given the dispersions. The hypothesis being tested is H_0: Inferred $-$ Contingent ≥ 0, therefore a high p-value ($\geq$90 per cent) indicates that Inferred $>$ Contingent, a low p-value ($\leq$10 per cent) indicates rejection of the null while an intermediate-sized p-value (10 per cent $< p$-value $<$ 90 per cent) indicates that WTP Inferred = WTP Contingent.

Tables 6 and 7 indicate that for high commitment costs, the IV method generates WTP values that are statistically significantly *greater* than the CV method. This finding applies equally to experienced and inexperienced subjects. However, for inexperienced subjects, in the majority of cases for which commitment costs are low, WTP values elicited with the IV method are *lower* than the CV method. Moreover, this finding is often the case when normative motivations are high rather than low. In many other cases,

16 We should note that in a follow-up experiment, we made some changes in the products under valuation by making the inferiority of the conventional product more salient. We do not report results from this experiment here due to space considerations but overall findings can by summarised like this: we still do not observe strong preference reversals but only weak evaluation mode effects. The IV method is even more susceptible to weak evaluation mode effects than the field experiment we report here. Details can be found in the working paper version (Stachtiaris, Drichoutis and Klonaris, 2011).

Table 6. Hypothesis tests for whether IV generates lower valuations than CV[a,*]: olive oil case

		Normative motivations			
		Price-sensitive		Price-insensitive	
Evaluation mode	Commitment cost	Low	High	Low	High
Panel A: experienced subjects					
More Joint	High	2.422 (0.998)	2.258 (0.995)	2.430 (0.991)	2.266 (0.985)
	Low	0.596 (0.743)	−1.394 (0.693)	0.604 (0.716)	0.440 (0.668)
Less Joint	High	1.420 (0.977)	1.256 (0.959)	1.428 (0.943)	1.264 (0.915)
	Low	−0.406 (0.306)	−0.570 (0.215)	−0.397 (0.337)	−0.562 (0.264)
More Isolated	High	1.551 (0.972)	1.387 (0.953)	1.560 (0.955)	1.396 (0.931)
	Low	−0.274 (0.385)	−0.439 (0.308)	−0.266 (0.396)	−0.430 (0.326)
Less Isolated	High	2.136 (0.998)	1.971 (0.997)	2.144 (0.993)	1.980 (0.990)
	Low	0.310 (0.638)	0.146 (0.574)	0.318 (0.630)	0.154 (0.570)
Panel B: inexperienced subjects					
More Joint	High	1.231 (0.980)	1.067 (0.940)	1.239 (0.939)	1.075 (0.890)
	Low	−0.595 (0.182)	−0.759 (0.117)	−0.587 (0.235)	−0.751 (0.176)
Less Joint	High	1.200 (0.994)	1.036 (0.960)	1.209 (0.955)	1.044 (0.903)
	Low	−0.626 (0.132)	−0.790 (0.076)	−0.617 (0.200)	−0.781 (0.144)
More Isolated	High	1.455 (0.999)	1.291 (0.985)	1.463 (0.985)	1.299 (0.949)
	Low	−0.371 (0.265)	−0.535 (0.192)	−0.363 (0.315)	−0.527 (0.250)
Less Isolated	High	1.558 (0.999)	1.394 (0.990)	1.567 (0.987)	1.403 (0.957)
	Low	−0.267 (0.328)	−0.431 (0.244)	−0.259 (0.369)	−0.423 (0.299)

Table C1 (Appendix C in supplementary data at *ERAE* online) shows the exact linear combination of coefficients being tested in each case. The numbers depicted without parentheses are the sum of the corresponding coefficients depicted in Table C1.

[a]The hypothesis tested is H_0: Inferred − Contingent ≥ 0. A rejection of the null is equivalent to H_1: Inferred $<$ Contingent.

*p-values in parentheses.

for example, when commitment costs are low and normative motivations are low, values elicited with IV are statistically indistinguishable than values elicited with CV.

Table 7. Hypothesis tests for whether IV generates lower valuations than CV[a,*]: Apple case

		Normative motivations			
		Price-sensitive		Price-insensitive	
Evaluation mode	Commitment cost	Low	High	Low	High
Panel A: experienced subjects					
More Joint	High	0.813 (0.994)	0.527 (0.959)	0.850 (0.987)	0.564 (0.937)
	Low	0.499 (0.923)	−0.101 (0.760)	0.536 (0.906)	0.250 (0.749)
Less Joint	High	0.578 (0.980)	0.292 (0.864)	0.615 (0.965)	0.329 (0.838)
	Low	0.264 (0.801)	−0.022 (0.467)	0.301 (0.794)	0.015 (0.518)
More Isolated	High	0.642 (0.991)	0.357 (0.909)	0.680 (0.983)	0.394 (0.888)
	Low	0.328 (0.857)	0.043 (0.562)	0.366 (0.849)	0.080 (0.595)
Less Isolated	High	1.025 (0.999)	0.739 (0.999)	1.062 (0.999)	0.777 (0.996)
	Low	0.711 (0.996)	0.425 (0.962)	0.748 (0.992)	0.463 (0.942)
Panel B: inexperienced subjects					
More Joint	High	−0.001 (0.498)	−0.287 (0.093)	0.036 (0.556)	−0.249 (0.195)
	Low	−0.315 (0.096)	−0.601 (0.004)	−0.278 (0.180)	−0.563 (0.031)
Less Joint	High	0.148 (0.830)	−0.138 (0.233)	0.185 (0.784)	−0.100 (0.354)
	Low	−0.166 (0.221)	−0.452 (0.013)	−0.129 (0.325)	−0.414 (0.072)
More Isolated	High	0.424 (0.997)	0.139 (0.749)	0.462 (0.970)	0.176 (0.727)
	Low	0.111 (0.692)	−0.175 (0.218)	0.148 (0.691)	−0.138 (0.328)
Less Isolated	High	0.340 (0.993)	0.054 (0.613)	0.378 (0.945)	0.092 (0.629)
	Low	0.026 (0.551)	−0.259 (0.100)	0.064 (0.588)	−0.222 (0.224)

Table C1 (Appendix C in supplementary data at *ERAE* online) shows the exact linear combination of coefficients being tested in each case. The numbers depicted without parentheses are the sum of the corresponding coefficients depicted in Table C1.

[a]The hypothesis tested is H_0: Inferred − Contingent ≥ 0. A rejection of the null is equivalent to H_1: Inferred $<$ Contingent.

*p-Values in parentheses.

In summary, we partially reconfirm Lusk and Norwood's (2009a, 2009b) results that found that for goods with high normative motivations, we should expect IV to generate lower valuations than CV. However, their model did not predict that for high commitment costs and low normative motivations, there

could be a case that IV generates higher valuations that CV. This calls for further scrutiny of the IV method with more diverse samples and goods.

5. Conclusions

We started this article with a series of questions which we are now ready to answer. In the field experiment we conducted, we could not replicate strong evaluation mode effects as in the original studies that scrutinised the sports cards market for preference reversals (List, 2002; Alevy, List and Adamowicz, 2011). In our experiment, we do not observe strong evaluation mode effects (i.e. preference reversals) with just one minor exception. However, we do observe some weak evaluation mode effects in the IV method, and thus it appears that IV is slightly more susceptible in generating inconsistent preference orderings than the CV method. Market experience did not make a difference to our results.

Our second aim was to re-examine the effectiveness of the IV method in mitigating social desirability bias. IV consistently generated higher valuations than CV in the case where commitment costs were high. On the other hand, with low commitment costs and high normative motivations, IV successfully mitigated social desirability by generating lower valuations than CV. This means that for people who buy a food quality product more often (low commitment costs) and believe the product encompasses normative (moral) considerations (high normative motivations), the IV method successfully mitigates social desirability bias and leads to lower valuations than the CV method. To the extent that hypothetical bias and social desirability bias were present in our study (and we have no reason to believe that our study would differ from other hypothetical studies), this is a sign that IV will work for certain products and consumers.

Further research on methods that mitigate social desirability bias and hypothetical bias are indeed warranted. It will take time and more studies of this kind to answer the question for which contexts, products and samples the IV method remains a promising method for mitigating biases. More research on this domain may help economists find ways to accurately measure values without having to develop a market for goods. All in all, we believe that this topic could indeed be a prime area for future economic research.

Supplementary Data

Supplementary data are available at *ERAE* online.

References

Aizaki, H., Nakashima, Y., Ujiie, K., Takeshita, H. and Tahara, K. (2011). Intention to access food risk information through internet-enabled mobile phones: the role of critical thinking. *Applied Economics Letters* 18: 1005–1009.

Alevy, J. E., List, J. and Adamowicz, W. (2011). How can behavioral economics inform non-market valuation? An example from the preference reversal literature. *Land Economics* 87: 365–381.

Alfnes, F. and Rickertsen, K. (2007). Extrapolating experimental-auction results using a stated choice survey. *European Review of Agricultural Economics* 34: 345–363.

Alfnes, F., Yue, C. and Jensen, H. H. (2010). Cognitive dissonance as a means of reducing hypothetical bias. *European Review of Agricultural Economics* 37: 147–163.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H. (1993). Report of the NOAA panel of contingent valuation. *Federal Register* 58: 4601–4614.

Becker, G. S. (1976). *The Economic Approach to Human Behavior*. Chicago, IL: University of Chicago Press.

Blaine, T. W., Lichtkoppler, F. R., Jones, K. R. and Zondag, R. H. (2005). An assessment of household willingness to pay for curbside recycling: a comparison of payment card and referendum approaches. *Journal of Environmental Management* 76: 15–22.

Boccaletti, S. and Nardella, M. (2000). Consumer willingness to pay for pesticide-free fresh fruit and vegetables in Italy. *International Food and Agribusiness Management Review* 3: 297–310.

Boyle, K. J., Bishop, R. C. and Welsh, M. P. (1985). Starting point bias in contingent valuation bidding games. *Land Economics* 61: 188–196.

Boyle, K. J., Johnson, F. R. and McCollum, D. W. (1997). Anchoring and adjustment in single-bounded, contingent-valuation questions. *American Journal of Agricultural Economics* 79: 1495–1500.

Brouwer, R., van Beukering, P. and Sultanian, E. (2008). The impact of the bird flu on public willingness to pay for the protection of migratory birds. *Ecological Economics* 64: 575–585.

Brugarolas, M., Martinez-Carrasco, L., Bernabeu, R. and Martinez-Poveda, A. (2009a). A contingent valuation analysis to determine profitability of establishing local organic wine markets in Spain. *Renewable Agriculture and Food Systems* 25: 1–19.

Brugarolas, M., Martínez-Carrasco, L., Martínez-Poveda, A. and Ruiz, J. J. (2009b). A competitive strategy for vegetable products: traditional varieties of tomato in the local market. *Spanish Journal of Agricultural Research* 7: 294–304.

Brummett, R. G. and Nayga, R. M. (2007). On the use of cheap talk in new product valuation. *Economics Bulletin* 2: 1–9.

Buzby, J. C., Fox, J. A., Ready, R. C. and Crutchfield, S. R. (1998). Measuring consumer benefits of food safety risk reductions. *Journal of Agricultural and Applied Economics* 30: 69–82.

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

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.

Corsi, A. (2007). Ambiguity of measured WTP for quality improvements when quantity is unconstrained: a note. *European Review of Agricultural Economics* 34: 501–515.

Crowne, D. P. and Marlowe, D. (1960). A new scale of social desirability independent of psychopathalogy. *Journal of Consulting Psychology* 24: 349–350.

Cummings, R. G. and Harrison, G. W. (1992). A new approach to eliciting incentive compatible values for public goods: the inference game. Mimeo, Department of Economics, University of South Carolina.

Donaldson, C., Thomas, R. and Torgerson, D. J. (1997). Validity of open-ended and payment scale approaches to eliciting willingness to pay. *Applied Economics* 29: 79–84.

Drichoutis, A. C., Lazaridis, P. and Nayga, R. M., Jr (2009). On consumers' valuation of nutrition information. *Bulletin of Economic Research* 61: 223–247.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research* 20: 303–315.

Frederick, S. and Fischhoff, B. (1998). Scope (in)sensitivity in elicited valuations. *Risk Decision and Policy* 3: 109–123.

Green, D., Jacowitz, K. E., Kahneman, D. and McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20: 85–116.

Grether, D. M. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69: 623–638.

Herriges, J. and Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management* 30: 112–131.

Holmes, T. P. and Kramer, R. A. (1995). An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. *Journal of Environmental Economics and Management* 29: 121–132.

Hsee, C. K. (1996). The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes* 46: 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S. and Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological Bulletin* 125: 576–591.

Hsu, J. L., Chen, H.-Y., Ming-Li Chen, C. and Chang, W.-H. (2009). Sustainability of fresh fruit certifications: willingness to pay measures. *Journal of Sustainable Agriculture* 33: 690–700.

Hu, W., Woods, T., Bastin, S., Cox, L. and You, W. (2011). Assessing consumer willingness to pay for value-added blueberry products using a payment card survey. *Journal of Agricultural and Applied Economics* 43: 243–258.

Johnson, E. J., Steffel, M. and Goldstein, D. G. (2005). Making better decisions: from measuring to constructing preferences. *Health Psychology* 24: S17–S22.

Kanninen, B. (1995). Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management* 28: 114–125.

Leggett, C. G., Kleckner, N. S., Boyle, K. J., Duffield, J. W. and Mitchell, R. C. (2003). Social desirability bias in contingent valuation surveys administered through in-person interviews. *Land Economics* 79: 561–575.

Lichtenstein, S. and Slovic, P. (2006). *The Construction of Preference*. Cambridge, UK: Cambridge University Press.

Lienhoop, N. and MacMillan, D. (2007). Valuing wilderness in Iceland: estimation of WTA and WTP using the market stall approach to contingent valuation. *Land Use Policy* 24: 289–295.

List, J., Berrens, R., Bohara, A. and Kerkvliet, J. (2004). Examining the role of social isolation on stated preferences. *American Economic Review of Agricultural Economics* 94: 741–752.

List, J. A. (2002). Preference reversals of a different kind: the 'more is less' phenomenon. *American Economic Review* 92: 1636–1643.

List, J. A. and Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values. *Environmental and Resource Economics* 20: 241–254.

Little, J. and Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. *Economics Bulletin* 3: 1–13.

Lusk, J. and Norwood, F. B. (2009a). Bridging the gap between laboratory experiments and naturally occurring markets: an inferred valuation method. *Journal of Environmental Economic and Management* 58: 236–250.

Lusk, J. and Norwood, F. B. (2009b). An inferred valuation method. *Land Economics* 85: 500–514.

Lusk, J., Pruitt, J. R. and Norwood, F. B. (2006). External validity of a framed field experiment. *Economics Letters* 93: 285–290.

Murphy, J. J., Stevens, T. H., Allen, P. G. and Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics* 30: 313–325.

Payne, J. W., Bettman, J. R. and Schkade, D. A. (1999). Measuring constructed preferences: towards a building code. *Journal of Risk and Uncertainty* 19: 243–270.

Plant, E. A., Devine, P. G. and Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: revisiting the fading and faking of racial prejudice. *Group Processes and Intergroup Relations* 6: 187–200.

Reaves, D. W., Kramer, R. A. and Holmes, T. P. (1999). Does question format matter? Valuing an endangered species. *Environmental and Resource Economics* 14: 365–383.

Ressurreição, A., Gibbons, J., Kaiser, M., Ponce Dentinho, T., Zarzycki, T., Bentley, C., Austen, M., Burdon, D., Atkins, J., Santos, R. S. and Edwards-Jones, G. (2012). Different cultures, different values: the role of cultural variation in public's WTP for marine species conservation. *Biological Conservation* 145: 148–159.

Rowe, R. D., Schulze, W. D. and Breffle, W. S. (1996). A test for payment card biases. *Journal of Environmental Economics and Management* 31: 178–185.

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys* 16: 621–655.

Simpson, S. N. and Hanna, B. G. (2010). Willingness to pay for a clear night sky: use of the contingent valuation method. *Applied Economics Letters* 17: 1095–1103.

Skuras, D. and Vakrou, A. (2002). Consumers' willingness to pay for origin labelled wine. *British Food Journal* 104: 898–912.

Slovic, P. (1995). The construction of preference. *American Psychologist* 50: 364–371.

Slovic, P. and Lichtenstein, S. (1983). Preference-reversals: a broader perspective. *American Economic Review* 73: 596–605.

Slovic, P. and Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology* 78: 1–18.

Smith, R. D. and Sach, T. H. (2010). Contingent valuation: what needs to be done. *Health Economics, Policy and Law* 5: 91–111.

Solomon, B. D. and Johnson, N. H. (2009). Valuing climate protection through willingness to pay for biomass ethanol. *Ecological Economics* 68: 2137–2144.

Stachtiaris, S., Drichoutis, A. C. and Klonaris, S. (2011). Preference reversals in contingent and inferred valuation methods. Munich Personal RePEc Archives Paper No. 32800. http://mpra.ub.uni-muenchen.de/32800/. Accessed in June 2012.

Thayler, M. (1981). Contingent valuation techniques for assessing environmental impacts: further evidence. *Journal of Environmental Economics and Management* 8: 27–44.

Zhongmin, X., Loomis, J., Zhiqiang, Z. and Hamamura, K. (2006). Evaluating the performance of different willingness to pay question formats for valuing environmental restoration in rural China. *Environment and Development Economics* 11: 585–601.