Fair farming: Preferences for fair labor certification using four elicitation methods

Drichoutis, Andreas C. and Vassilopoulos, Achilleas and Lusk, Jayson and Nayga, Rodolfo M.

Agricultural University of Athens, Oklahoma State University, University of Arkansas

3 March 2015

Fair farming: Preferences for fair labor certification using four elicitation methods

Andreas C. Drichoutis[*1], Achilleas Vassilopoulos [†1], Jayson L. Lusk[‡2], and Rodolfo M. Nayga, Jr.[§3]

[1]Agricultural University of Athens
[2]Oklahoma State University
[3]University of Arkansas & Norwegian Agricultural Economics Research Institute & Korea University

Abstract: High profile cases of exploitative labor practices have increased concerns over agricultural working conditions. However, it is unclear whether and to what extent the public is willing to trade-off fair working conditions for higher prices and food imports. We implement a large-scale survey to uncover Greek consumer preferences for a food labeling system that certifies fair working conditions for the workers employed at all production stages of agricultural production. Empirical findings from several disciplines suggest that results from contingent valuation surveys can be susceptible to hypothetical bias, social desirability bias, and lack of consequentiality. To test these issues, we use the 'cheap talk' method (Kling et al., 2012), Lusk and Norwood's (2009b) Inferred Valuation (IV) method and the consequentiality scripts employed in Vossler and Evans (2009) and Vossler and Watson (2013). We also test predictions of reference dependent theory by testing whether framing the valuation question as an 'Equivalent Loss' (EL) differs from classical 'Willingness-to-pay elicitation' (WTP). We collected responses from more than 3,800 consumers in the cities of Athens and Ioannina in Greece. Our results show that neither the cheap talk nor the consequentiality script had any effect on elicited valuations. In contrast, the IV method

[*]Lecturer, Department of Agricultural Economics & Rural Development, Agricultural University of Athens, Iera Odos 75, 11855, Greece, e-mail: adrihout@aua.gr.

[†]PhD candidate, Department of Agricultural Economics & Rural Development, Agricultural University of Athens, Iera Odos 75, 11855, Athens, Greece, e-mail: avas@aua.gr.

[‡]Regents Professor and Willard Sparks Endowed Chair, Department of Agricultural Economics, Oklahoma State University, Stillwater, OK 74078, USA, e-mail: jayson.lusk@okstate.edu.

[§]Professor and Tyson Endowed Chair, Department of Agricultural Economics & Agribusiness, Division of Agriculture, University of Arkansas, Fayetteville, AR 72701, USA, and Adjunct Professor, Norwegian Agricultural Economics Research Institute and Korea University, e-mail: rnayga@uark.edu.

appears to mitigate social desirability bias. We also find that values elicited under WTP are larger than values elicited under EL, which rejects neoclassical preferences. When social desirability is taken out of our estimates, we find that consumers are willing to pay an average premium of 72 cents/Kg for strawberries with fair labor certification, which is equivalent to 49% of current market prices.

Keywords: fair labor label; willingness to pay; equivalent loss; contingent valuation; inferred valuation; consequentiality; cheap talk; uncertainty scale.

JEL Classification Numbers: C83; C93; D12; Q13.

1 Introduction

Recently, there has been a surge of labor exploitation incidents in the Greek agricultural farm sector. The most recent incident in April 2013 involved 33 Bangladeshi workers being shot and injured by their supervisors at a strawberry farm as they protested for being unpaid for several months. This incident brought mass media attention to the long-term issue of labor exploitation in the agricultural farming sector. Subsequently, a steady stream of cases of mistreatment and unfair working conditions in farm businesses have emerged, resulting in public outcry in Greece and the European Union. Public anger was manifested by a strong (albeit temporary) decline in the demand for strawberries. Despite the public backlash, farmers have argued that in order to produce at prices that consumers are willing to pay, and to be competitive with imported agricultural products, they cannot afford to provide generous benefits for their employees (e.g., minimum wage, maximum working hours, sick leave, housing). Without resorting to illegal employment practices, producers fear they will be unable to profitably operate.

This paper seeks to contribute to the debate regarding the controversy about the trade-offs between fair working conditions and the competitiveness of local agricultural products. We use non-market valuation techniques designed to uncover the underlying preferences of Greek consumers towards a food labeling system that certifies "fair" working conditions

for the workers employed at all stages of agricultural production. Our aim is to understand whether consumers' alleged disapproval of unfair working practices is reflected in their willingness-to-pay (WTP) a premium above regular prices of conventional agricultural products. Most relevant to our work is a labeling system that is already in place in the global agro-food system known as Fair Trade (FT) labels.[1] However, fair labor labeling is distinct from fair trade labeling in that the later is mainly focused on commodities or products which are typically exported from developing to developed countries and thus, is focused on helping producers in *developing countries.*

Although previous research has studied ethically-related food marketing claims related to animal welfare, environment, or fair trade, we are not aware of prior research that has focused specifically on consumer preferences for agricultural employee pay and working conditions. However, recent works by Howard and Allen (2006, 2010) and Hustvedt and Bernard (2010) provide accumulated evidence in favor of labor-related information on a variety of products. Using both hypothetical and non-hypothetical methods, the studies found that consumers are willing to pay a premium for products, the production of which mandates subsistence wages to all employees in the production chain. Our aim in this paper is to take this line of research a step forward by introducing and testing labels that ensure not only subsistence wages but also govern working hours, access to decent housing, and personal hygiene facilities as well as health care services. Given that employment in Greek farms is mostly seasonal and reliant on illegal immigrants who might have less recourse through traditional legal channels, these additional conditions are likely aspects of an agro-food labor market that consumers might find desirable.

To uncover consumers' preferences for a fair labor certification system, this study uses traditional stated preferences methods augmented with recent methodological advances in the field designed to identify and weed out potential biases. We conducted a large scale questionnaire based experiment in two cities of Greece, Athens and Ioannina. We collected data from personal interviews of more than 3,800 subjects. In addition to the empirical objective related to fair labor labels, we also explore several methodological issues that are relevant to nonmarket valuation, such as social desirability bias, hypothetical bias, consequentiality of the survey, and certainty of responses. The next section describes the survey-experiment and the relevant methodological issues we address.

---

[1]For a recent critical overview of the economic theory behind Fair Trade, see Dragusanu et al. (2014).

2 The valuation survey-experiment and methodological issues

The Contingent Valuation (CV) method has become one of the most popular methods to measure WTP values for public and private goods, services, or amenities. Although it was principally developed in environmental and transport economics, it has made considerable headway in the valuation of food products over the last decades (e.g., Buzby et al., 1998; Corsi, 2007). Most, if not all, CV studies are conducted in hypothetical contexts, particularly in environmental valuation studies where a real market with salient payments is difficult to establish (Carson, 2012; Haab et al., 2013; Kling et al., 2012).

The CVM involves creating a hypothetical valuation scenario in which consumers are asked to state their WTP for the product in question. Empirical findings from several disciplines suggest that results from the CVM can be susceptible to social desirability bias. There are now several studies that show that when people are asked to predict other people's value, as in the Inferred Valuation method (discussed momentarily), they state a different value than their own (e.g., Frederick, 2012; Kurt and Inman, 2013; Loewenstein and Adler, 1995; van Boven et al., 2000, 2003). Lusk and Norwood (2009b) used a modification of Levitt and List's (2007) additive utility model to illustrate how social desirability may carry over to non-market valuation and produce inflated bids that misrepresent respondents' preferences. Their so called Inferred Valuation (IV) method, addresses social desirability bias by asking respondents to state their beliefs about the average consumer's valuation for a good.

Lusk and Norwood (2009a,b) argued that the IV method generates valuations that are less likely to suffer from normative or moral response biases (such as social desirability bias), and they found that responses to the IV method better predicted actual shopping behavior than did those from a CV method. They also found that the IV method produced less hypothetical bias when social desirability was present. 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. In the spirit of Lusk and Norwood (2009a,b), Pronin (2007) argues that people tend to recognize biases in human judgment except when biases are their own, which implies that predictions over other peoples' preferences should mitigate biases.

Ever since the National Oceanographic and Atmospheric Administration (NOAA) (Arrow et al., 1993) criticized the open-ended format as providing 'erratic and biased' responses, the favored elicitation format in the CVM literature has been the dichotomous choice (DC) format. Although other alternatives that were not considered by the NOAA panel were later developed, Carson and Groves (2007) offer a typology of elicitation formats that shows

that all formats can be seen as generalizations of the DC format. The DC format has been favored due to its well known property of incentive compatibility. This is due to the Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975) which states that for the case of more than two alternatives (i.e., non-DC formats), no non-dictatorial strategy-proof voting procedure exists.[2]

This is to say that any response format with at least three possible outcomes is subject to individual manipulation (i.e., it is not incentive compatible). This does not imply that any binary DC format is incentive compatible but that, by elimination, only a DC format could be incentive compatible assuming subjects believe that their response is consequential (meaning there is some probability that the respondent's answer will actually influence the provision of the good).[3] As we discuss in the next paragraph, the consequentiality of the survey is a key condition in the assumption of incentive compatibility, and yet it is an issue that has only lately recieved much attention.

Much of the early literature has evolved around the presumption that hypothetical bias prevails in CV estimates (i.e., that people tend to state a higher WTP in hypothetical questions as opposed to non-hypothetical questions). As such, a number of approaches have been proposed to reduce or elimate hypothetical bias. For example, the 'cheap talk' method has been used to potentially reduce hypothetical bias by reminding participants of the tendency among people to inflate their (hypothetical) valuations (Kling et al., 2012). As discussed above, the issue of consequentiality has only been given attention in the last few years. Carson and Groves (2007) argue that to produce useful information about an agent's preferences, survey results must be seen as potentially consequential. Under consequentiality, survey respondents are explicitly told that their responses to preference questions will influence provision of the good under valuation. As a result of Carson and Groves's (2007) suggestion, the literature examining the merits of consequential surveys has been rapidly growing (Carson et al., 2014; Herriges et al., 2010; List and Price, 2013; Mitani and Flores, 2013; Poe and Vossler, 2011; Vossler et al., 2012; Vossler and Evans, 2009; Vossler and Watson, 2013).

An additional challenge we face with this survey-experiment is the issue of reference-dependent preferences. In the presence of information asymmetry, reference points may

---

[2]The theorem was formalized by Gibbard (1973) and Satterthwaite (1975) and noted in passing by Dummett and Farquharson (1961). See also Svensson and Reffgen (2014).

[3]While proponents of the DC format take this result about the incentive compatibility of the DC format as granted for any type of good, Carson et al. (1997) show that the DC format is not incentive compatible in the case of provision of a new private or quasi-public good. The incentive compatibility of the DC format can be restored for quasi-public or private goods only if the binary choice is between two different forms of the good, so that the valuation question represents a change in the good (Carson et al., 2001; Carson and Groves, 2007).

be formulated differently for each consumer and thus, heterogeneity in valuations may not only reflect taste heterogeneity. In our context, for example, for the share of consumers aware of labor exploitation in farms, the current endowment corresponds to products whose production line entails undesirable practices. As such, the availability of certified alternatives is seen as an opportunity to upgrade the (extrinsic) quality of goods they consume, at a cost equal to the price premium of such alternatives. Clearly, this premium is better framed as their willingness-to-pay (WTP), which is defined as the maximum amount a consumer would be willing to pay in return for an increase (or upgrade) in his consumption for a good. This is the valuation measure used in most studies eliciting homegrown values (e.g., Bateman et al., 2000). On the other hand, there are uninformed consumers who think that currently available goods are produced using fair labor practices and thus, according to their (wrong) perception, the introduction of a certification system would induce an extra-cost for goods which are the same with the ones they are already consuming. As a result, their valuation is better framed as an 'Equivalent loss' (EL), defined as the maximum amount a consumer would be willing to pay in place of a reduction (or downgrade) in her consumption of a good. Under Hicksian preferences, these two valuation measures should be equal to each other (EL=WTP), since ordering of consumption bundles is independent of individual's endowment. However, if preferences are formed as in Tversky and Kahneman's (1992) reference-dependent theory, this distinction is important. For this reason and to better approximate the average welfare of fair labor labels we use both WTP and EL questions in a within-subjects basis.

Aside the empirical necessity for the use of both measures, their relative magnitude from a methodological point of view is very interesting as well. In two competing hypothesis regarding reference-dependent preferences, Munro and Sugden (2003), show how the WTA/WTP gap is caused both by loss aversion in the good and by loss aversion in money. This hypothesis is also verified by the results of Bateman et al. (1997) using experimental markets and Bateman et al. (2000) using the CV method. On the other hand, Kahneman et al. (1990) provide evidence of Tversky and Kahneman's (1991) original 'no loss in buying' (NLIB) hypothesis which states that buyers do not value the money they give up in a transaction as a loss; that is money spent is not 'coded' as a loss. This is also confirmed in Novemsky and Kahneman (2005) who firmed up the NLIB hypothesis by proposing a theory of the conditions under which the gains and losses associated with a transaction are mentally integrated prior to evaluation, rather than being evaluated separately. They propose that when a loss and an equal and opposite gain are integrated in this way, painful perceptions of loss do not arise. Because consumers normally have budget reserves, that is, reserves of money that are available for unanticipated spending, when faced an opportunity to buy a good and when they are able to finance this spending from their budget reserve, gains and

6

losses are integrated. As a result, the money that has to be spent to buy the good is already seen as a token for unspecified goods and thus money outlays are not perceived as losses. In contrast, if the individual faces an unanticipated buying opportunity which she can finance only by forgoing some specific consumption plan, the act of buying involves a definite loss, separable from the gain; and so the money payment is perceived as a loss. A subsequent adversarial collaboration paper of Bateman et al. (2005) however, provides evidence that the cost of buying is indeed subject to loss aversion.

In our specific application normalizing the utility of gains to zero, if preferences are reference dependent in the form of Munro and Sugden (2003), then we should observe EL≥WTP (for example, see Bateman et al., 2000, for a proof), irrespective of whether downgrading in EL is perceived as a loss or not.[4] On the other hand, if the NLIB hypothesis is true, then there are three different competing scenarios about the relative magnitude of WTP and EL. In the first scenario when money spent to avoid the downgrade is covered by budget reserves (as in WTP) and thus is not perceived as a loss, EL is greater or equal to WTP (EL≥WTP) notwithstanding the loss perceptions in EL treatments. In the other two scenarios, the cost of avoiding the downgrade as well as the downgrade *per se* are perceived as losses in EL and as such we expect EL≥ (≤)WTP when losses in the money (good) domain are weighted more heavily than equi-util losses in the other dimension. In conclusion, while EL>WTP is indicative of reference-dependent preferences with or without NLIB, EL<WTP is supportive of the NLIB hypothesis. We test for these competing hypotheses in our survey by framing the valuation questions either as a WTP or as an EL. Next section describes the experimental design we adopt to test the hypotheses described above.

3   Experimental design and questionnaire development

To elicit valuations for the fair labor certification system, we chose a pack of 500gr of strawberries as our product of interest. This was our chosen product for a number of reasons. First, we needed a product which is widely consumed and would appeal to most consumers. Second, we wanted a product that can be sold in packages that could carry a labeling certification system; strawberries can and are often sold in a packaged form. Third, the production process of strawberries is a (manual) labor intensive process (labor contributes to 45-50% of total cost of strawberries (Poinssot, 2013)).

All subjects were first informed about the fair labor certification system. Because there

---

[4]Normalizing the utility of gains to zero assumes that there is no "extra kick" from gaining a good, other than its consumption utility. This is the original and the most commonly used formulation of prospect theory for riskless choices.

was the chance that subjects were familiar with a 'fair trade' label that could confound their perception of a 'fair labor' label, we wanted to make sure that all subjects would be actually valuing a 'fair labor' label by providing them a script with relevant information for fair labor labeling (the script can be found in Appendix A).

To answer the methodological issues we raised in the previous section, we adopted a design with elements of a within- as well as a between-subjects design (Charness et al., 2012). For the between-subjects design we adopted a 4×2 design where we vary the scripts accompanying the valuation questions (control (no scripts) vs. cheap talk vs. consequentiality vs. cheap talk & consequentiality) as well as the elicited valuation measure (equivalent loss vs. willingness to pay). The five bid amounts used for the Discrete Choice format (20 cents vs. 40 cents vs. 70 cents vs. 100 cents vs. 120 cents) were selected based on projected historical prices of strawberries as well as feedback we received from the pilot survey (see Appendix C for a discussion of selection of bid amounts). The bid amounts were varied on a between subject basis so that each subject was asked for his/her valuation at only one bid amount. Table 1 shows our experimental design and the per treatment number of subjects. Valuations were elicited using the CV method as well as the IV method on a within subjects basis. The order was counterbalanced.

Table 1: Experimental design

		Equivalent Loss					Willingness to pay					
	Bid amount	20	40	70	100	120	20	40	70	100	120	Total
IV then CV	Control	46	47	47	47	47	51	50	49	49	48	481
	Consequentiality	44	47	47	47	47	48	47	48	48	48	471
	Cheap talk	47	47	47	47	47	48	47	47	47	48	472
	Cheap talk & Consequentiality	47	47	47	47	47	49	48	48	48	48	476
CV then IV	Control	48	48	48	49	46	50	49	49	49	49	485
	Consequentiality	46	47	47	46	48	50	49	49	50	49	481
	Cheap talk	46	47	47	47	47	49	49	49	49	49	479
	Cheap talk & Consequentiality	47	47	47	48	48	49	48	48	49	49	480
	Total	371	377	377	378	377	394	387	387	389	388	3825

Notes: CV stands for Contingent valuation; IV stands for Inferred Valuation.

The cheap talk script was compiled from several sources as well as our own previous work (e.g., Bulte et al., 2005; Lusk, 2003) and reads as follows:

"In a minute you'll be asked whether you are willing to pay a certain amount for strawberries.

8

This question will be hypothetical, that is, you will not actually have to pay. In general, people experience difficulties in answering hypothetical questions. They often state they are willing to pay an amount larger than the amount they are willing to pay in reality.

One reason why this happens is because when the time comes to actually make the payment, they also consider that this money won't be available for other purchases. Therefore, when the question is hypothetical, it is easier to exaggerate their response.

Before answering the willingness to pay question, try to think whether you are really willing to pay this amount for strawberries and that this amount will no longer be available for other purchases."

The consequentiality script was adopted from Vossler and Watson (2013) and Vossler and Evans (2009) and reads as follows:

"We would like to inform you that the survey results will become available to producers, traders and retailers of agricultural products as well as to the wider general public of consumers. This means that this survey could affect the decision of producers, traders and retailers to adopt a Fair Labor certification system for strawberries as well as the average price of strawberries."

After the script(s) was(were) read, the valuation questions followed. In the control (no script) treatment, the valuation treatments followed right after information about the fair labor label was given. The willingness to pay valuation measure was framed as:

"Assume you are given a pack of half a kilo of conventional strawberries without any certification [show picture 2]. Would you be willing to pay XX cents so that you can exchange it with a similar pack of strawberries certified with a fair labor label [show picture 3]?"[5]

Conversely, the equivalent loss measure was framed as:

"Assume you are given a pack of half a kilo of strawberries certified with a fair labor label [show picture 3]. Would you be willing to pay XX cents so that you can avoid exchanging it with a pack of conventional strawberries without any certification? [show picture 2]"

Following the literature on certainty scales (Champ et al., 1997; Morrison and Brown, 2009), every CV discrete choice question was followed by a question asking respondents to

---

[5]The pictures can be found in Appendix B.

state how certain they were about their response on a 10-point scale anchored with the labels "Not certain at all" and "Very certain" .

Similar to the CV questions, the IV questions were formatted to elicit the willingness to pay measure:

> "Assume that an average consumer is given a pack of half a kilo of conventional strawberries without any certification [show picture 2]. Do you think s/he would be willing to pay XX cents so that s/he can exchange it with a similar pack of strawberries certified with a fair labor label [show picture 3]?"

or the equivalent loss valuation measure:

> "Assume that an average consumer is given a pack of half a kilo of strawberries certified with a fair labor label [show picture 3]. Do you think s/he would be willing to pay XX cents so that s/he can avoid exchanging it with a pack of conventional strawberries without any certification? [show picture 2]"

A consequentiality question was also included to allow us to test for differences between respondents with different consequentiality perceptions of the survey. The question was adopted from Vossler and Watson (2013) and Vossler et al. (2012). Participants had to indicate the indirect consequences of the survey on a five point likert scale anchored by "not at all" and "very much". The question was framed as:

> "To what extent do you believe that your answers in this survey will be taken into account by producers, traders and retailers?"

The questionnaire also elicited respondents' beliefs about the likelihood of hypothetical bias and social desirability bias (using the social desirability scale (SDS) of Stöber (2001)). Recently, research in the area of non-market valuation has identified political affiliation and means of provision of the good i.e., privately or collectively, as a determinant of willingness to pay for public goods (e.g., Dupont and Bateman, 2012). This mainly originates from a body of research that has found significant differences in the degree of concern for environmental goods expressed by individuals with different political affiliations. While our valuation exercise does not concern a public good, aspects of the fair labor label may appeal to political ideology e.g., the minimum wage provision. Therefore, a political ideology question was included which we adopted from the European Social Survey. A set of demographic questions on age, gender, education, household size and perceived income position of the household was also asked as well as questions related to respondents' price sensitivity with respect to grocery shopping and purchase frequency of strawberries.

4 Data collection methods

A pilot questionnaire was pre-tested in February-March 2014 in the city of Athens with 160 subjects, after which several adjustments were made. The full scale survey was then launched on April 1, 2014 and questionnaires were filled in until June 11, 2014. The period of data collection was mainly dictated by the fact that strawberries are mainly traded during April, May and early June. Consumers were randomly intercepted in front of the main entrance of various supermarkets. In all, 11,510 subjects were intercepted in the cities of Athens and Ioannina and 3,825 agreed to take part in the survey resulting in a cooperation rate of 33.23%. Of course, several subjects walked out during an interview or opted not to respond to certain questions, which further reduced the available number of subjects for statistical analysis.

Regarding the demographic profile of our sample, Table 2 shows that the vast majority of respondents were females (66.36%). This is not as problematic as it may seem given that the primary shoppers in Greece, as in many other countries, are typically females. For example, one study estimates that 75% of principal household shoppers in the US are females (Mediamark Research and Intelligence, 2009). Therefore, the gender composition of our sample is not representative of the population of the two cities but it might better represent the grocery shopping population. Since we also asked respondents to report on the age and gender composition of their household, we can also compare the demographic profile of respondents' households with that of the 2001 census (which is the latest available census for which basic demographic information are available). The comparison (see Table 2) shows that discrepancies with the 2001 census are small.

Table 2: Comparison of gender and age groups between survey respondents, their household members and the 2001 census (percentages)

	Males	0-9	10-19	20-29	30-39	40-49	50-59	60-69	≥70
Respondents	33.64	0.00	3.47	24.07	22.37	23.50	17.29	7.13	2.18
Households	48.24	8.22	11.89	20.77	15.39	16.61	16.96	7.05	3.03
Census	48.45	9.22	11.46	16.37	16.06	14.54	11.93	10.45	9.96

5 Data analysis and results

As in every experiment, it is critical to explore whether randomization to treatment was successful by testing if the observable characteristics are balanced across the between subjects treatments. Table 3 shows the basic descriptive statistics for a set of observable

characteristics. With respect to demographic variables, there is no significant difference between treatments for gender (Pearson's $\chi^2 = 5.132$, p-value=0.644), education (Pearson's $\chi^2 = 20.168$, p-value=0.979), income (Pearson's $\chi^2 = 39.326$, p-value=0.076), age (Kruskal-Wallis $\chi^2 = 5.599$, p-value=0.587) and household size (Kruskal-Wallis $\chi^2 = 3.341$, p-value=0.852).

We also find no difference between treatments with respect to the Social Desirability score (Kruskal-Wallis $\chi^2 = 6.128$, p-value=0.525), political ideology (Kruskal-Wallis $\chi^2 = 8.036$, p-value=0.329), purchase frequency of strawberries (Pearson's $\chi^2 = 30.800$, p-value=0.671), price sensitivity in purchasing decisions (Pearson's $\chi^2 = 22.302$, p-value=0.767), perceived likelihood of hypothetical bias for the respondent (Pearson's $\chi^2 = 31.952$, p-value=0.276) and perceived likelihood of hypothetical bias for other respondents (Pearson's $\chi^2 = 23.605$, p-value=0.702). Thus, we can safely conclude that the randomization to treatment on a wide set of observable characteristics was successful.

5.1   Descriptive data analysis

Before we proceed with the econometric analysis we can gain interesting insights by looking at the raw data. Figure 1 graphs the percentage of Yes/No responses elicited with the CV method in the discrete choice question for each of the four script treatments. Two observations are noteworthy. First, it appears that in all treatments, the elicited values are responsive to prices (bid amounts); we observe a decline of 'yes' responses when the bid amount increases. The 40 cents bid yields statistically significantly lower proportion of 'Yes' responses as compared to the 20 cent bid (p-value<0.001); the 70 cent bid yields significantly lower responses than the 40 cent bid (p-value=0.001); the 100 cent bid is also lower and statistically different than the 70 cent bid (p-value=0.015) while the 120 cent bid does not differ with respect to the 100 cent bid (p-value=0.829). In addition, a significant percentage of subjects states a high value for the fair labor certification label even when the bid is up to 120 cents. This is interesting since the amount of 120 cents represents a 171% increase in the price of strawberries (with respect to a price of 70 cents/500 gr).

Second, it appears that the different scripts are not effective in changing response patterns with respect to the control (no script) treatment. This is confirmed by proportion tests where we find that the proportion of 'Yes' responses in the Control treatment is not significantly different from the Consequentiality script treatment (p-value=0.225); or the Cheap talk script treatment (p-value=0.720); or the Cheap talk & Consequentiality script treatment (p-value=0.082).

Figure 2 graphs the proportion of responses in the discrete choice question for the IV

Table 3: Descriptives statistics of subjects' observable characteristics

Variable name and description	Variable levels	N	Mean	S.D.
Gender: Male dummy		3763	0.34	0.47
Age		3721	39.97	13.77
Hsize: Household size		3708	3.40	1.25
Shopper: Dummy for major grocery shopper		3754	0.78	0.41
SDS: Social Desirability Scale		3636	11.35	2.68
Political: Political ideology scale		3312	4.66	2.14
Educ: Education level	Up to primary school	3708	3.61%	
	Up to Junior Hi-school		4.75%	
	Up to Senior Hi-school		22.92%	
	Some college or university student		21.90%	
	University graduate		37.30%	
	Post-graduate studies		9.52%	
Income: Household's economic position	Bad or Very bad	3702	5.73%	
	Below average		11.16%	
	Average		47.41%	
	Above average		19.75%	
	Good or Very good		15.96%	
HBiasOwn: Perceived likelihood of hypothetical bias for the respondent	Not likely at all	3725	40.70%	
	Unlikely		31.03%	
	Neither likely, nor unlikely		15.09%	
	Likely		11.49%	
	Very likely		1.69%	
HBiasOther: Perceived likelihood of hypothetical bias for other respondents	Not likely at all	3709	4.26%	
	Unlikely		20.22%	
	Neither likely, nor unlikely		32.27%	
	Likely		33.67%	
	Very likely		9.57%	
PurchFreq: Purchase frequency (when strawberries are available)	Never	3707	14.32%	
	Once a month		25.06%	
	2-3 times a month		21.12%	
	Once a week		28.16%	
	2-3 times a week		9.55%	
	More than 2-3 times a week		1.78%	
PriceSens: Price sensitivity	Not important at all	3695	1.14%	
	Not important		3.82%	
	Neither important, nor unimportant		15.21%	
	Important		42.14%	
	Very important		37.70%	

Notes: S.D. stands for standard deviation. For variables with levels, the percentage of each level is denoted instead of the mean. For each variable the sample was restricted to observations for which subjects gave responses to both the CV and IV methods.
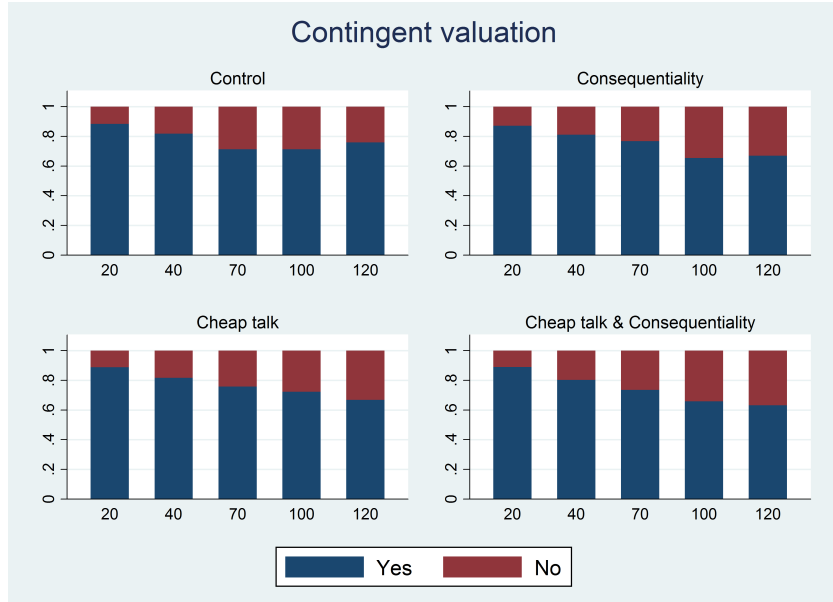
Figure 1: Percentage of Yes/No responses in the CVM per treatment

method. The findings are similar to those in the CV method: we find no effect of the scripts on responses (p-value=0.285, 0.413 and 0.866 when we compare the Control treatment with the Cheap talk script, Consequentiality script and Cheap talk & Consequentiality script treatments, respectively). In addition, while proportion of 'Yes' responses declines for the first two bid amounts, it is flat after 70 cents. Proportion tests indicate that the 40 cents bid amount yields a significantly lower proportion of 'Yes' responses when compared with the 20 cents bid (p-value<0.001), while the 70 cents bid yields lower 'Yes' proportion than the 40 cents bid (p-value<0.001). However, the 100 cents and 120 cents do not yield different responses when compared with the 70 cents bid (p-value=0.443 and 0.672, respectively).

What is even more interesting is that the IV method seems to work as advertised. Figure 2 shows that for all bid amounts, the IV method yields a significantly lower proportion of 'Yes' responses when compared with CV method shown in Figure 1. Proportion tests show that differences are highly significant for all bid amounts (p-values<0.001 in all cases). The difference in the proportion of 'Yes' is larger than 30% for all bid amounts. This is an indication that the IV method was effective in mitigating social desirability and hypothetical bias. Thus, valuations from the IV method may be more valid for predicting market outcomes.

Figures 3 and 4 compare the elicited value measures (Equivalent loss vs. Willingness to pay) for the CV and IV methods, respectively. It is obvious that since the proportion of 'Yes' responses is higher for WTP than EL, then it is implied that values elicited under WTP are larger than values elicited under EL. Proportions tests confirm the insight that the proportion of 'Yes' for WTP is larger than EL under the CV method (p-value=0.017)
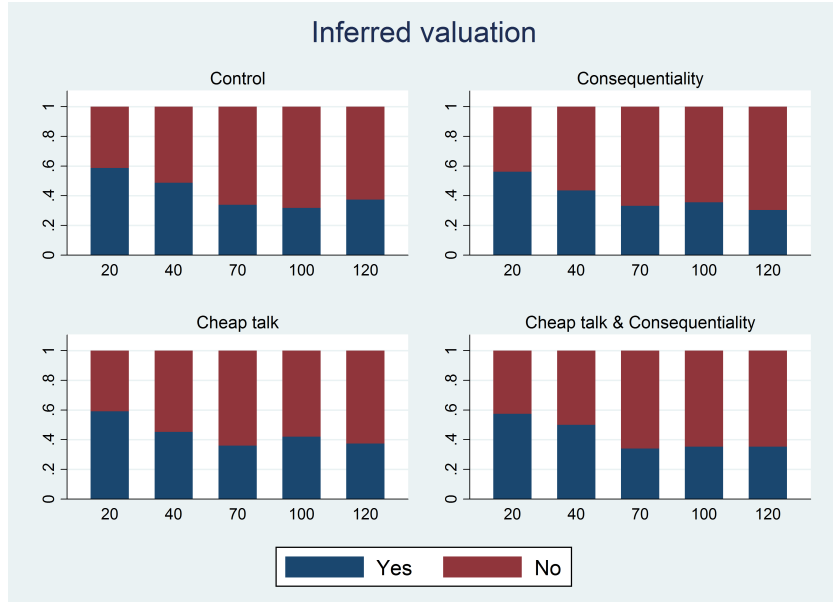
Figure 2: Percentage of Yes/No responses in the IVM per treatment

and under the IV method (p-value<0.001).

5.2 Econometric analysis

To check whether the results obtained above hold under conditional analysis, we estimated interval regression models with clustered standard errors at the individual level to account for the fact that each person provided responses under both CV and IV methods. Table 4 shows coefficient estimates from several specifications. Model (1) is our basic specification where only the treatment variables are included in the regression. We confirm that none of the scripts (Cheap talk, Consequentiality or their combination) has any effect on elicited valuations. On the other hand, the IV dummy is highly significant indicating that subjects under the IV method are stating lower valuations, which is a likely indication that this method successfully mitigates social desirability and hypothetical bias. The WTP dummy is also highly significant indicating that elicited valuations are higher when the valuation question is framed as a WTP question rather than as an EL question. Our result supports Tversky and Kahneman's (1991) 'no loss in buying' hypothesis which states that buyers do not value the money they give up in a transaction as a loss. Finally, there is evidence of significant order effects indicating that when the IV question was asked first, then subjects tended to align their response in the CV question. This could be due to the fact that answering the IV valuation question first and the CV question second, made subjects think of their own biases.
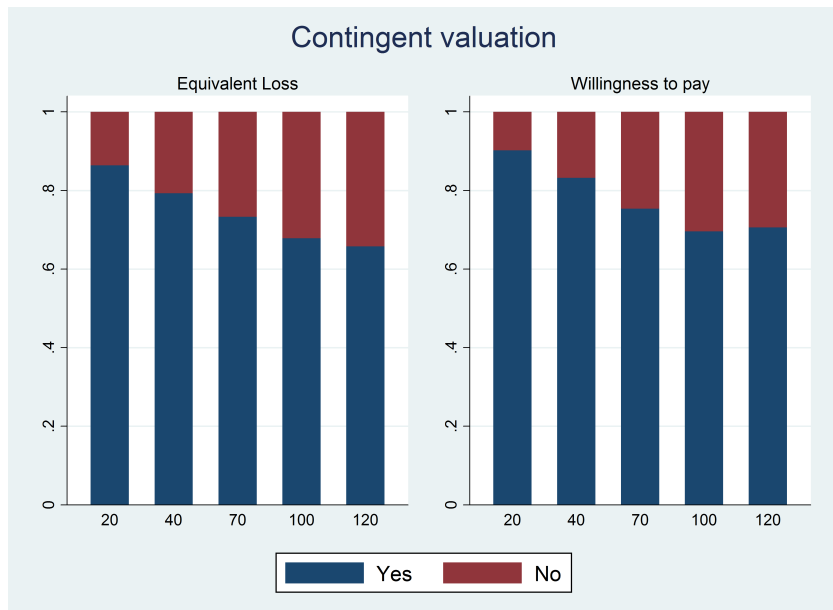
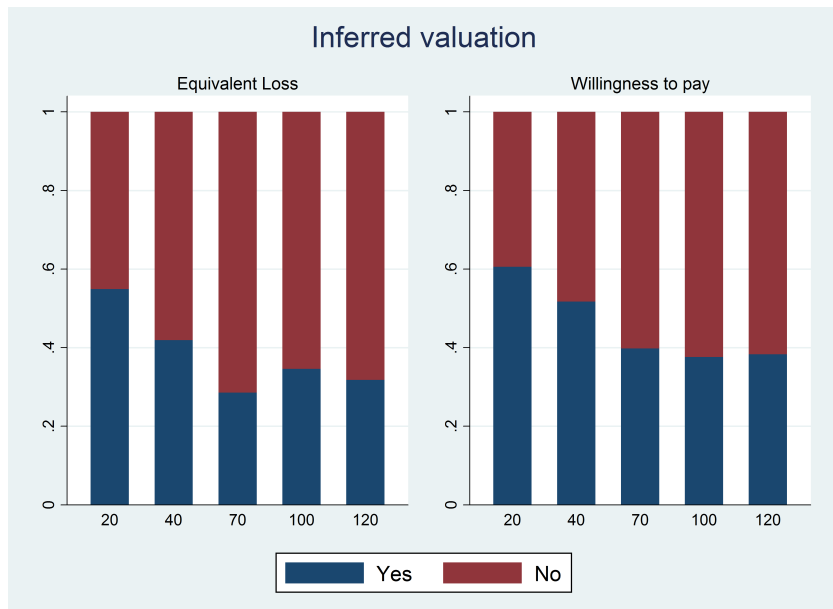Figure 3: Percentage of Yes/No responses in the CVM per value measure

Figure 4: Percentage of Yes/No responses in the IVM per value measure

Table 4: Interval regression estimates

	Without certainty correction			With certainty correction		
	Full sample	Consequential	Incosequential	Full sample	Consequential	Incosequential
	(1)	(2)	(3)	(4)	(5)	(6)
Cheap talk	-10.435	-4.026	-18.803*	-1.576	-6.070	-2.078
	(7.976)	(13.602)	(9.624)	(9.003)	(15.722)	(10.499)
Consequentiality	3.198	25.630*	-12.628	1.574	14.411	-8.915
	(8.008)	(13.981)	(9.664)	(9.093)	(15.959)	(10.667)
Cheap talk & Consequentiality	-6.995	9.162	-17.832*	-10.824	-9.897	-11.897
	(7.980)	(13.779)	(9.702)	(9.103)	(15.839)	(10.716)
IV	-153.459***	-192.484***	-131.080***	-134.251***	-167.648***	-114.248***
	(12.278)	(25.748)	(12.853)	(12.316)	(26.192)	(12.622)
WTP	25.224***	42.251***	11.736*	23.559***	32.195***	15.376**
	(6.002)	(11.262)	(6.865)	(6.753)	(12.285)	(7.658)
Order	-34.028***	-46.103***	-30.629***	-36.137***	-53.404***	-29.555***
	(6.177)	(11.427)	(7.206)	(7.083)	(13.855)	(7.948)
Constant	197.542***	256.788***	160.681***	153.451***	220.106***	112.295***
	(12.053)	(27.148)	(11.885)	(10.835)	(26.629)	(10.338)
$ln(\sigma)$	5.091***	5.309***	4.887***	5.205***	5.449***	4.968***
	(0.077)	(0.130)	(0.094)	(0.088)	(0.152)	(0.105)
N	7544	4060	3484	7544	4060	3484
AIC	8998.514	4589.512	4189.506	9619.611	5105.709	4312.026
BIC	9053.943	4639.983	4238.754	9675.039	5156.181	4361.273

Clustered standard errors in parentheses * $p<0.1$, ** $p<0.05$, *** $p<0.01$

As we have discussed in previous sections, recent literature has argued that consequentiality is needed for incentive compatibility. We therefore explore differences in results for people who believe and do not believe their answers are consequential. Based on the five point Likert consequentiality question, we divided subjects into 'consequential' and 'inconsequential'. We defined the consequential sample as involving subjects that stated that they believed their answers will be taken into account by producers, traders and retailers on a medium or higher likelihood (3 or more on the 5-point Likert scale). The inconsequential sample was defined as those that perceived that the survey would have little or no effect to producers, traders and retailers.

Columns (2) and (3) in Table 4 show estimates from the consequential and inconsequential subsamples, respectively. The results are roughly similar with those from the full sample with one exception. A few script treatment variables are now statistically significant (albeit only at the 10% level). However, this not is not likely a robust result. Table A1 in Appendix D shows estimates of the consequential and inconsequential subsamples with the addition of demographic and attitudinal variables in the list of independent variables. As evident, the script treatment variables are no longer statistically significant. Thus, we conclude that there are no differential effects with respect to the script treatment variables between the subsamples.

A final piece of robustness check for our results comes from certainty corrections. Following previous literature (Morrison and Brown, 2009), we recoded the 'Yes' responses in the CV question using a rating of 7 in the certainty question as the cut-off. Columns (4), (5) and (6) in Table 4 present coefficients for the full sample as well as the consequential and inconsequential subsamples, respectively. None of our conclusions changes. The only difference we observe with the addition of demographics (estimates are shown in Table A1 in the Appendix D) is that for the inconsequential subsample, framing of the valuation question does not matter i.e., we observe that willingness to pay equals equivalent loss.

We can then proceed in graphing the aggregate demand curves for strawberries with fair labor certification. Figure 5 does exactly that for valuations elicited with CV and IV. Given that IV mitigates social desirability bias, it would be advisable for producers and retailers to base their marketing decisions on the IV curve. Each point on this curve indicates the percentage of respondents that would buy half a kilo of strawberries with fair labor certification at the premium projected on the Y axis. The fact that the IV curve intersects the X axis indicates that a percentage of consumers would not buy strawberries with fair labor certification even when these are offered at no premium at all.
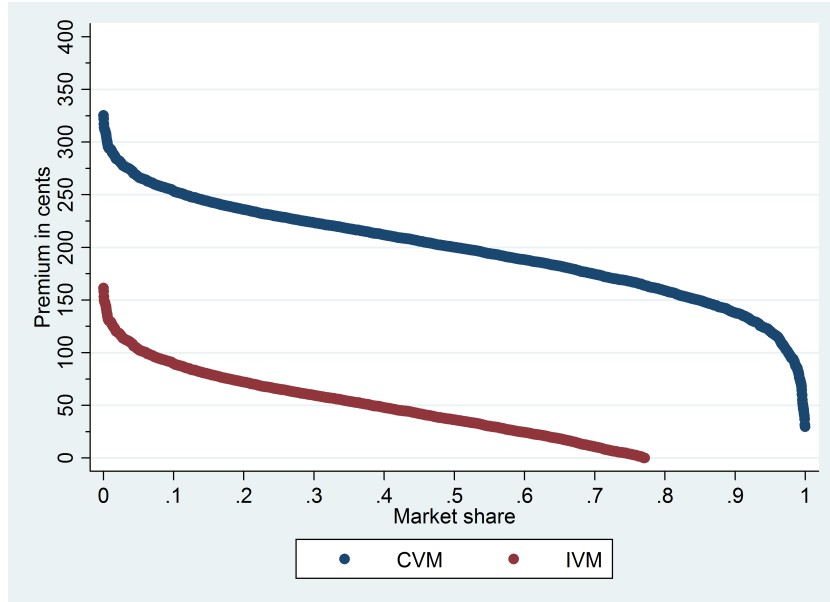
Figure 5: Aggregate demand curves for the CV and IV elicitation methods

6   Conclusions and discussion

There have been increasing calls around the world for urgent action to tackle widespread abuse of migrant workers in the agri-food sector. This paper sought to contribute to the debate regarding the controversy about the trade-offs between fair working conditions and the competitiveness of local agricultural products. To do so, we used a Contingent Valuation survey designed to uncover the underlying preferences of Greek consumers towards a food labeling system that certifies fair working conditions for the workers employed at all production stages of agricultural products. In addition to this important empirical objective, we also examined several methodological issues i.e, how we can mitigate hypothetical bias, social desirability bias and lack of consequentiality. We also examined predictions of reference dependent theory by testing whether framing the valuation question as an 'Equivalent Loss' (EL) differs from classical 'Willingness-to-pay' elicitation.

While the IV method worked as advertised and was successful in mitigating social desirability bias, none of the scripts we employed (i.e., cheap talk, consequentiality and their combination) had any effect on elicited valuations. This is surprising given that there is now a well established literature which has identified cheap talk scripts as effective in mitigating hypothetical bias (e.g., Lusk, 2003; Morrison and Brown, 2009; Özdemir et al., 2009). Similarly, consequentiality has emerged as an important addition to standard CV studies (Vossler and Evans, 2009; Vossler and Watson, 2013). One way to interpret our non-result from the evaluation of these two scripts is that using scripts from previous studies is not a

19

panacea that will work in every single study. More research is needed to further test the effectiveness of the cheap talk and consequentiality scripts in reducing hypothetical bias in different contexts.

Our results also show that values elicited under WTP are larger than values elicited under EL, which supports Tversky and Kahneman's (1991) 'no loss in buying' hypothesis which states that buyers do not value the money they give up in a transaction as a loss. When social desirability is taken out of our estimates, we find that consumers are willing to pay an average premium of 72 cents/Kg of strawberries with fair labor certification, which is equivalent to 49% of current market prices. This can be valuable information for producers and retailers that seek to differentiate their products and are wondering whether costs associated with product differentiation can be recouped from potential customers. All in all, our study provides the first economic estimates in the literature of premiums associated with fair labor certification and shows that while fair labor certification is a road not yet taken, it might carry significant benefits for crop laborers, producers and consumers. The positive premiums that our respondents are willing to pay suggest that people are aware of fair labor issues and value the better treatment of workers in the agricultural sector.

References

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

Bateman, I., D. Kahneman, A. Munro, C. Starmer, and R. Sugden (2005). Testing competing models of loss aversion: an adversarial collaboration. *Journal of Public Economics 89*(8), 1561–1580.

Bateman, I., A. Munro, B. Rhodes, C. Starmer, and R. Sugden (1997). A test of the theory of reference-dependent preferences. *The Quarterly Journal of Economics 112*(2), 479–505.

Bateman, I. J., I. H. Langford, A. Munro, C. Starmer, and R. Sugden (2000). Estimating four hicksian welfare measures for a public good: A contingent valuation investigation. *Land Economics 76*(3), 355–373.

Bulte, E., S. Gerking, J. A. List, and A. de Zeeuw (2005). The effect of varying the causes of environmental problems on stated wtp values: evidence from a field study. *Journal of Environmental Economics and Management 49*(2), 330–342.

Buzby, J. C., J. A. Fox, R. C. Ready, and S. R. Crutchfield (1998). Measuring consumer benefits of food safety risk reductions. *Journal of Agricultural and Applied Economics 30*(1), 69–82.

Carson, R. T. (2012). Contingent valuation: A practical alternative when prices aren't available. *Journal of Economic Perspectives 26*(4), 27–42.

Carson, R. T., N. E. Flores, and N. F. Meade (2001). Contingent valuation: Controversies and evidence. *Environmental and Resource Economics 19*(2), 173–210.

Carson, R. T. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics 37*(1), 181–210.

Carson, R. T., T. Groves, and J. A. List (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists 1*(1), 171–207.

Carson, R. T., T. Groves, and M. J. Machina (1997). Stated preference questions: Context and optimal response. *Paper presented at the National science foundation preference elicitation symposium, University of California, Berkeley*.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum (1997). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management 33*(2), 151–162.

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

Corsi, A. (2007). Ambiguity of measured wtp for quality improvements when quantity is unconstrained: a note. *European Review of Agricultural Economics 34*(4), 501–515.

Dragusanu, R., D. Giovannucci, and N. Nunn (2014). The economics of fair trade. *Journal of Economic Perspectives 28*(3), 217236.

Dummett, M. and R. Farquharson (1961). Stability in voting. *Econometrica 29*(1), 33–43.

Dupont, D. P. and I. J. Bateman (2012). Political affiliation and willingness to pay: An examination of the nature of benefits and means of provision. *Ecological Economics 75*(0), 43–51.

Frederick, S. (2012). Overestimating others willingness to pay. *Journal of Consumer Research 39*(1), 1–21.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica 41*(4), 587–601.

Haab, T. C., M. G. Interis, D. R. Petrolia, and J. C. Whitehead (2013). From hopeless to curious? Thoughts on Hausman's dubious to hopeless critique of contingent valuation. *Applied Economic Perspectives and Policy 35*(4), 593–612.

Herriges, J., C. Kling, C.-C. Liu, and J. Tobias (2010). What are the consequences of consequentiality? *Journal of Environmental Economics and Management 59*(1), 67–81.

Howard, P. H. and P. Allen (2006). Beyond organic: Consumer interest in new labelling schemes in the central coast of California. *International Journal of Consumer Studies 30*(5), 439–451.

Howard, P. H. and P. Allen (2010). Beyond organic and fair trade? An analysis of ecolabel preferences in the united states. *Rural Sociology 75*(2), 244–269.

Hustvedt, G. and J. C. Bernard (2010). Effects of social responsibility labelling and brand on willingness to pay for apparel. *International Journal of Consumer Studies 34*(6), 619–626.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990). Experimental tests of the endowment effect and the Coase theorem. *98*(6), 1325–1348.

Kling, C. L., D. J. Phaneuf, and J. Zhao (2012). From exxon to bp: Has some number become better than no number? *Journal of Economic Perspectives 26*(4), 3–26.

Kurt, D. and J. J. Inman (2013). Mispredicting others' valuations: Self-other difference in the context of endowment. *Journal of Consumer Research 40*(1), 78–89.

Levitt, S. D. and J. A. List (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives 21*(2), 153–174.

List, J. A. and M. K. Price (2013). Using field experiments in environmental and resource economics. *NBER Working Paper No. 19289*.

Loewenstein, G. and D. Adler (1995). A bias in the prediction of tastes. *The Economic Journal 105*(431), 929–937.

Lusk, J. and B. F. Norwood (2009a). Bridging the gap between laboratory experiments and naturally occurring markets:an inferred valuation method. *Journal of Environmental Economic and Management 58*(2), 236–250.

Lusk, J. L. (2003). Effects of cheap talk on consumer willingness-to-pay for golden rice. *American Journal of Agricultural Economics 85*(4), 840–856.

Lusk, J. L. and B. F. Norwood (2009b). An inferred valuation method. *Land Economics 85*(3), 500–514.

Mediamark Research and Intelligence (2009). Despite decades of gains in the workforce, women still the predominant household shoppers. http://www.gfkmri.com/PDF/MRIPR_111209_HouseholdShoppers.pdf. Last accessed on December 10, 2014.

Mitani, Y. and N. E. Flores (2013). Hypothetical bias reconsidered: Payment and provision uncertainties in a threshold provision mechanism. *Environmental and Resource Economics*, 1–22.

Morrison, M. and T. C. Brown (2009). Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies. *Environmental and Resource Economics 44*(3), 307–326.

Munro, A. and R. Sugden (2003). On the theory of reference-dependent preferences. *Journal of Economic Behavior & Organization 50*(4), 407–428.

Novemsky, N. and D. Kahneman (2005). The boundaries of loss aversion. *Journal of Marketing Research 42*(2), 119–128.

Özdemir, S., F. R. Johnson, and A. B. Hauber (2009). Hypothetical bias, cheap talk, and stated willingness to pay for health care. *Journal of Health Economics 28*(4), 894–901. 10.1016/j.jhealeco.2009.04.004.

Poe, G. L. and C. A. Vossler (2011). Consequentiality and contingent values: An emerging paradigm. The international handbook on non-market environmental valuation. Cheltenham, UK: Edward Elgar Publishing.

Poinssot, A. (2013, June). The migrant workers trapped in slave-like conditions in greece. http://www.mediapart.fr/journal/international/230613/migrant-workers-trapped-slave-conditions-greece. Last accessed on December 10, 2014.

Pronin, E. (2007). Perception and misperception of bias in human judgment. *Trends in Cognitive Sciences 11*(1), 37–43.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory 10*(2), 187–217.

Stöber, J. (2001). The social desirability scale-17 (sds-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment 17*(3), 222–232.

Svensson, L.-G. and A. Reffgen (2014). The proof of the Gibbard-Satterthwaite theorem revisited. *Journal of Mathematical Economics 55*(0), 11–14.

Tversky, A. and D. Kahneman (1991). Loss aversion in riskless choice: A reference-dependent model. *106*(4), 1039–1061.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

van Boven, L., D. Dunning, and G. Loewenstein (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology 79*(1), 66–76.

van Boven, L., G. Loewenstein, and D. Dunning (2003). Mispredicting the endowment effect: Underestimation of owners selling prices by buyers agents. *Journal of Economic Behavior & Organization 51*(3), 351–365.

Vossler, C. A., M. Doyon, and D. Rondeau (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics 4*(4), 145–71.

Vossler, C. A. and M. F. Evans (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. *Journal of Environmental Economics and Management 58*(3), 338–345.

Vossler, C. A. and S. B. Watson (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization 86*, 137 147.

A   Questionnaire scripts

Information for the fair labor informational script was compiled from three relevant to the fair labor label websites: an auditing non-profit organization that formally measures and accredits best employment practices (http://www.fairworkingconditions.ie), the fair labor association (http://www.fairlabor.org) and the Food Alliance certification programme (http://foodalliance.org). The script read like this:

> "The Fair Labor label [show picture 1][6] can be certified by various organizations like the Fair Working Conditions.ie which is a non-profit international organization that aims in recognizing and improving employment conditions. Such a label ensures that the product is produced in a business farm that strictly follows the standards set by the International Labour Organisation (ILO). These standards concern the maximum number of working hours per week, the legal compensation and labor benefits required by state law for each activity sector as well as workers' hygiene conditions in their working place. In addition, the standards prohibit child labor and bind the employer for no discrimination on the basis of race, nationality, or any other criteria."

B   Pictures

---

[6]The picture can be found in Appendix B.

Picture 1

Picture 2

Picture 3

C Selection of bid amounts

To select the bid amounts for the discrete choice valuation question, we first examined historical data for strawberries from the Athens Central Market Organization S.A. which is one of the major markets for vegetable and meat wholesalers in Greece.[7] The historical data show that the bulk of strawberry sales takes place between April and May (smaller quantities of strawberries are also traded in February, March and June) which also determined the data collection period of our survey-experiment. A simple linear projection of strawberry prices with data from April 2008 to May 2013, projected an average mode price for April-May 2014 at €1.458/Kg. This is remarkably close to the realized average mode price for April-May 2014 at €1.465/Kg. Given that our valuation product was selected to be a pack of 500gr of strawberries (corresponding projected price is €0.73/500 gr) and that the valuation question concerned an upgrade (or downgrade, depending on the elicited valuation measure), we decided that our price range would center around €0.70 taking also into account feedback from a pilot survey. Therefore, the full price range was selected as: 20 cents, 40 cents, 70 cents, 100 cents and 120 cents.

D Additional tables

---

[7]The Athens Central Market Organization S.A. (http://www.okaa.gr/) is a public organization supervised by the Ministry of Rural Development and Food. It serves as an organized host of about 600 vegetable and meat wholesalers and operates in an area of about 26 hectares.

Table A1: Interval regression estimates (with demographics)

	Without certainty correction			With certainty correction		
	Full sample	Consequential	Incosequential	Full sample	Consequential	Incosequential
	(1)	(2)	(3)	(4)	(5)	(6)
Cheap talk	-10.004	-5.011	-17.604*	0.821	-5.425	1.569
	(8.860)	(15.423)	(10.702)	(9.798)	(17.018)	(11.609)
Consequentiality	10.541	24.280	-0.598	11.285	20.808	1.586
	(8.969)	(15.901)	(10.878)	(10.024)	(17.543)	(11.966)
Cheap talk & Consequentiality	-3.782	10.789	-14.183	-7.098	-3.562	-10.973
	(8.732)	(15.578)	(10.588)	(9.755)	(16.952)	(11.698)
IV	-163.861***	-205.998***	-142.361***	-140.506***	-172.517***	-122.545***
	(14.479)	(31.089)	(15.618)	(13.978)	(29.054)	(14.975)
WTP	23.825***	41.896***	11.616	18.416**	27.435**	10.848
	(6.693)	(12.883)	(7.751)	(7.263)	(12.997)	(8.493)
Order	-36.990***	-49.069***	-34.703***	-37.587***	-55.205***	-31.051***
	(6.997)	(13.214)	(8.305)	(7.789)	(15.114)	(8.940)
Age	-0.948***	-0.955*	-0.902***	-1.121***	-1.096*	-1.156***
	(0.294)	(0.518)	(0.350)	(0.331)	(0.569)	(0.400)
Gender	-4.828	0.318	-0.649	15.971**	28.206*	14.860
	(7.031)	(12.669)	(8.217)	(7.886)	(14.641)	(9.122)
$Educ_2$	17.138	18.658	16.325	14.099	12.989	11.823
	(24.906)	(42.467)	(31.642)	(28.015)	(46.094)	(37.424)
$Educ_3$	22.888	45.385	20.961	44.864*	57.529	50.681
	(20.094)	(35.509)	(25.000)	(23.264)	(39.315)	(30.831)
$Educ_4$	-2.374	1.097	2.225	12.215	10.420	22.907
	(20.756)	(35.807)	(25.761)	(23.849)	(39.362)	(31.442)
$Educ_5$	10.546	13.154	18.559	34.418	25.665	51.631*

	(19.871)	(34.460)	(24.845)	(23.050)	(37.975)	(30.769)
Educ_6	-10.343	-36.998	14.466	13.599	-5.048	35.946
	(21.892)	(38.326)	(26.859)	(25.070)	(41.656)	(32.556)
Hsize	-1.189	-4.211	-0.449	-4.157	-8.670*	-2.570
	(2.642)	(4.549)	(3.333)	(2.933)	(5.165)	(3.635)
Shopper	-1.481	0.915	-4.325	14.531	18.820	11.093
	(8.195)	(14.532)	(9.920)	(9.280)	(16.515)	(10.983)
Political	-0.130	3.786	-4.183**	-0.676	0.624	-3.369
	(1.485)	(2.582)	(1.903)	(1.649)	(2.818)	(2.077)
SDS	8.564***	3.314	9.657***	8.715***	3.663	9.626***
	(1.480)	(2.452)	(1.790)	(1.655)	(2.680)	(1.962)
HBiasOhter_2	41.399**	64.628**	20.406	36.005*	43.424	28.834
	(18.235)	(32.145)	(22.881)	(19.814)	(33.933)	(24.948)
HBiasOther_3	33.055*	58.042*	14.626	35.856*	44.954	31.412
	(17.669)	(31.311)	(22.280)	(19.396)	(33.467)	(24.416)
HBiasOther_4	29.447*	57.144*	12.700	28.698	43.656	24.022
	(17.713)	(31.582)	(22.148)	(19.375)	(33.774)	(24.152)
HBiasOther_5	9.804	-1.102	18.791	14.273	-0.802	32.416
	(19.516)	(34.341)	(24.331)	(21.541)	(37.612)	(26.784)
HBiasOwn_2	-5.541	-13.870	2.963	-26.309***	-34.250**	-19.393*
	(7.758)	(13.870)	(9.184)	(8.863)	(16.034)	(10.173)
HBiasOwn_3	8.750	-14.284	19.246	-30.684***	-62.211***	-16.515
	(9.863)	(16.892)	(12.584)	(11.551)	(21.505)	(13.974)
HBiasOwn_4	2.014	-10.565	4.007	-32.871**	-51.727**	-29.075*
	(11.196)	(18.890)	(14.243)	(12.773)	(22.386)	(16.142)
HBiasOwn_5	17.096	65.579	-1.705	-30.925	-28.955	-31.736
	(26.693)	(54.842)	(28.976)	(30.060)	(57.151)	(33.699)
PurchFreq_2	66.967***	58.716***	62.158***	57.650***	42.351*	54.762***
	(11.784)	(21.651)	(13.129)	(12.889)	(23.244)	(14.322)
PurchFreq_3	57.546***	48.908**	48.179***	64.985***	63.732***	47.580***
	(11.773)	(21.250)	(13.327)	(13.473)	(24.425)	(14.641)
PurchFreq_4	66.355***	66.010***	49.550***	82.640***	87.036***	59.126***

	(1)	(2)	(3)	(4)	(5)	(6)
	(11.825)	(21.826)	(12.960)	(13.899)	(25.617)	(14.526)
PurchFreq$_5$	88.770***	91.399***	58.107***	106.687***	100.539***	81.619***
	(15.408)	(26.859)	(18.588)	(17.666)	(29.910)	(20.791)
PurchFreq$_6$	94.388***	96.559**	75.980*	107.096***	131.162***	49.077
	(28.220)	(44.987)	(39.386)	(30.715)	(49.688)	(40.203)
PriceSens$_2$	8.923	21.046	-9.842	29.847	-9.308	20.800
	(32.211)	(71.787)	(33.943)	(34.888)	(71.897)	(36.145)
PriceSens$_3$	3.648	0.217	-6.608	21.922	-62.093	39.871
	(28.887)	(67.031)	(29.036)	(31.215)	(66.855)	(31.160)
PriceSens$_4$	-9.928	-2.609	-26.281	11.562	-61.979	24.230
	(28.181)	(66.145)	(27.967)	(30.397)	(65.830)	(29.716)
PriceSens$_5$	-23.441	-25.563	-32.436	1.074	-72.295	15.345
	(28.448)	(66.414)	(28.386)	(30.576)	(66.249)	(29.964)
Athens	-4.480	5.315	-6.847	1.416	9.175	1.716
	(6.906)	(12.287)	(8.323)	(7.605)	(13.456)	(9.039)
Constant	71.128	156.537	71.715	-11.753	185.207*	-59.218
	(46.364)	(95.490)	(52.442)	(51.182)	(99.767)	(58.645)
$ln(\sigma)$	5.105***	5.344***	4.902***	5.195***	5.431***	4.972***
	(0.086)	(0.147)	(0.106)	(0.096)	(0.164)	(0.117)
N	6346	3416	2930	6346	3416	2930
AIC	7405.604	3863.316	3437.730	7889.676	4252.824	3532.210
BIC	7655.561	4090.356	3659.092	8139.632	4479.864	3753.573

Clustered standard errors in parentheses * p<0.1, ** p<0.05, *** p<0.01