On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments

Tiziana de-Magistris*, Azucena Gracia, and Rodolfo M. Nayga, Jr.

We test whether the use of an honesty priming task can help mitigate hypothetical bias in stated preference choice experiments (CE). Using a between-sample design, we conducted experiments with seven treatments. Our results suggest that marginal willingness to pay estimates from hypothetical CE with an honesty priming task are not significantly different from marginal valuations from non-hypothetical CE. Values from both of these treatments are lower than those from three other hypothetical treatments, while values from the three non-hypothetical treatments are not significantly different from each other.

Key words: Honesty priming, hypothetical bias, willingness to pay.

JEL codes: C18, C23, C90, D12.

Eliciting people's preferences for various goods using stated preference methods is a common practice in the applied economics and marketing literature. One of the methods used in stated preference elicitation is choice experiment (CE). Arguably, the CE approach is now the most widely used method in valuing products or attributes. However, a well-known shortcoming of the stated preference CE approach is hypothetical bias, defined as the difference between values obtained through hypothetical methods and the values (or what an individual might actually pay for the provision of the good) obtained through non-hypothetical methods. It is well documented that individuals overstate their willingness to pay (WTP) in hypothetical settings. In fact, a number of empirical papers have measured the magnitude of this hypothetical bias[1]. However, in spite of the abundance of studies about hypothetical bias, there is no definitive consensus about why people give different WTP values in hypothetical and non-hypothetical settings (Harrison 2006; Loomis 2011), or about how to calibrate experiments to obtain real WTP values (Murphy et al. 2005).

Harrison (2006) stated that one of the sources of hypothetical bias is that a good is not actually paid for or delivered in hypothetical settings; in other words, respondents do not have an economic incentive to reveal their true WTP values. Consequently, several researchers in the CE literature started using the so-called non-hypothetical or real choice experiment (RCE), which incorporates both an incentive compatible mechanism and real products to overcome the two limitations stated by Harrison (2006). For example, a RCE was used in Carlsson and Martinsson (2001), Cameron et al. (2002), Lusk and Schroeder (2004), Ding et al. (2005), Alfnes et al. (2006), Lusk et al. (2008), ~~Johansson-Stenman and Svedsäter (2008)~~, Chang et al. (2009), Loomis et al. (2009), ~~Volinsky et al. (2009), and Yue and Tong (2009)~~ to compare results from hypothetical and real choice experiments. With the exception of Carlsson and Martinsson (2001), who did not find any difference between the

*Tiziana de Magistris is a Marie Curie Fellow and research economist, and Azucena Gracia is a senior sesearch economist at the Unidad de Economía Agroalimentaria y de los Recursos Naturales, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Gobierno de Aragón, Zaragoza (Spain). Rodolfo M. Nayga, Jr. is a professor and holds the Tyson Endowed Chair at the Department of Agricultural Economics and Agribusiness, Division of Agriculture, University of Arkansas, Fayetteville, AR, and is also an adjunct professor at the Department of Food and Resource Economics, Korea University. The study has been funded by the European Union Marie Curie Project FOODLABELS_PIOF-GA-2009-253323 entitled "Do consumers value food labels? An assessment of the impact of information and personality traits on the demand for food labels." The authors would like to thank Brian Roe, Riccardo Scarpa, Kiyokazu Ujiie, and the anonymous journal reviewers for their helpful suggestions. Any remaining errors are the responsibility of the authors.

[1] For example, List and Gallet (2001) and Murphy et al. (2005) conducted meta-analyses of experimental studies to measure the magnitude of the hypothetical bias.

© The Author (2013). Published by Oxford University Press on behalf of the Agricultural and Applied Economics Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

hypothetical and non-hypothetical CE,[2] the rest of the papers provided strong evidence suggesting that results from hypothetical CE are different from those using RCE. Moreover, Chang et al. (2009) also found that non-hypothetical choices are a better approximation of true preferences than hypothetical choices based on a comparison not only of hypothetical CE with RCE, but also with actual market shares. The interpretation of these findings is that WTP values from RCE can be assumed to be the true values corresponding to actual payments in the marketplace (Chang et al. 2009).

However, it is sometimes difficult or impossible to conduct a RCE for several reasons. First, one needs the actual products to be able to properly conduct a RCE. Ideally, this means that a researcher must possess all the product profiles presented in the choice sets; this can be challenging given that many product concepts that researchers want to test with CE are not yet available on the market or have not yet been fully developed. Second, the RCE can be expensive and time-consuming to implement since subjects have to be paid a participation fee and actual transactions have to be made during the experiment.

Due to these difficulties, various ex-ante calibration methods have been applied in hypothetical experiments to mitigate the hypothetical bias. One of the most commonly used ex-ante calibration methods is the cheap talk script introduced in the seminal paper by Cummings and Taylor (1999). The cheap talk script explains the problem of hypothetical bias to the respondents prior to the administration of the valuation question. However, there has been no consensus on the effects of cheap talk, not only in the CE literature but also in the broader preference elicitation literature. For example, Lusk (2003) found that cheap talk did not reduce WTP values of knowledgeable consumers. He also reported that estimated WTP calculated from hypothetical responses with cheap talk was not significantly lower than willingness to pay estimates from hypothetical responses without cheap talk. Moreover, Brummett, Nayga, and Wu (2007) pointed out that their cheap talk script was not able to remove the hypothetical bias because there were no differences in their WTP estimates with or without cheap talk. On the other hand, Tonsor and Shupp (2011) reported that cheap talk in CEs conducted online could reduce the absolute value of the mean WTP, while Silva et al. (2011) found that their cheap talk script eliminated the hypothetical bias in a retail setting.

On the other hand, from a social psychology perspective, a cheap talk script could provide persuasive information with a social context that can make people behave in a desired way through communication (Jacquemet et al. 2011). Hence, the cheap talk script can be considered a form of manipulating chosen words that can be used to make respondents behave in the desired way to reveal their true preferences. However, Joule et al. (2008) stated that persuasive information is a necessary but not sufficient condition to automatically trigger proper behavior. According to these authors, information and persuasion can help change knowledge, modify attitudes, and induce genuine awareness, but a gap usually exists between "good ideas" and actions. Indeed, the study conducted by Jacquement et al. (2013) confirmed that the cheap talk script could have no effect on triggering sincere bidding. These authors proposed a new and alternative *ex ante* technique taken from social psychology called the "solemn oath"; they argued that a solemn oath can be used as a truth-telling commitment device by asking bidders to swear on their honor to provide honest answers prior to participating in a second-price auction. Their results suggest that the solemn oath improves the revelation of true preferences in both real and hypothetical auctions.

Given the inconsistency of findings on the effectiveness of cheap talk and the initial positive results on the use of the solemn oath in reducing hypothetical bias, it seems that approaches based on eliciting honest answers might be a promising area to further investigate vis-à-vis the cheap talk script. The theoretical foundation of this argument is based on the induced value theory (Smith 1976), which states that three conditions must be satisfied to solicit incentive-compatible behavior: monotonicity, salience, and dominance. Among these conditions, the most relevant criticism of hypothetical CE is the lack of the salience condition due to the absence of rewards directly related to the decisions that the subject makes during the experiment. Hypothetical bias may result when respondents do not take the hypothetical task seriously, or when they do not exert sufficient cognitive effort to provide accurate answers due to lack of economic incentives. In

[2] They used a within-subject non-hypothetical choice experiment.

other words, overstating WTP values could be driven by the tendency of respondents to lie in hypothetical settings.

As mentioned above, the findings of Jacquemet et al. (2011) regarding the solemn oath are promising. However, given that a solemn oath involves participants making a commitment, the possibility exists that some subjects may be bothered by this "heavy handedness." It is also possible that oath-taking may not be taken seriously by certain people for a variety of reasons (e.g., cultural background). We propose a new type of *ex ante* calibration in the same spirit as the honesty pledge of Jacquement et al. (2011; 2013) for eliciting "honest" answers. We also borrow this approach, termed "honesty priming," from the social psychology literature. Our proposed approach is based on the automatic activation of mental representation (i.e., honesty) proposed in the auto-motive model by Bargh (1990), rather than asking respondents to activate their own honesty by committing themselves to do so. According to Bargh (1997), automatic thought processes involve reflexive responses to certain triggering conditions. These processes require only that a stimulus event or object be detected by an individual's sensory system. Once that triggering event is detected, the process runs to completion without an individual's awareness. It is well demonstrated in social psychology that "priming" can unconsciously influence peoples' perception, evaluations, behavior, and choice (Maxwell, Nye, and Maxwell 1999; Bargh et al. 2001; Kay and Ross 2003; Chartrand et al. 2008). In other words, when people are incidentally exposed to some cues or words in an unrelated task, these stimuli can activate different buying goals, thereby influencing their subsequent decisions in a non-conscious manner (Chartrand et al. 2008). For example, Maxwell, Nye, and Maxwell (1999) demonstrated that participants who were primed for fairness showed more cooperative behavior, and consequently had a more positive attitude towards the seller. Bargh et al. (2001) also pointed out that when participants were primed with the concept of automatic achievement, the goal of performing better was activated in an unrelated subsequent task. Similarly, Kay and Ross (2003) demonstrated a high correlation between people given the cooperative and competitive priming condition and their deliberative intention to cooperate and compete, respectively.

Recent findings from papers in experimental economics that include economic incentive also support priming as an empirical tool to make salient some stereotypes or social identities that usually influence people's real-world behavior. To illustrate, Dee (2009) showed that priming an athletic identity concept to college student-athletes contributes to their academic underperformance. Benjamin et al. (2009) suggested that making ethnic, racial, and gender category norms salient affects economic preferences. Moreover, Benjamin et al. (2012) found that by manipulating religious identity, Protestants increased contributions to public goods. Similarly, Mckay et al. (2010) found that priming religious concepts strongly increased the costly punishment of unfair behavior for subjects who had previously donated to a religious organization. Finally, Drouvelis et al. (2010) illustrated that the use of priming techniques activates the concept of cooperation in a social dilemma game by increasing contributions levels.

In this study, we attempt to respond to four questions. Firstly, does honesty priming mitigate the hypothetical bias in hypothetical CEs? Second, if we find that honesty priming indeed reduces hypothetical bias, how much is it reduced? In particular, we wish to know if the hypothetical bias is totally reduced. Since cheap talk is another mechanism[3] more commonly used to potentially reduce hypothetical bias, our third question is: Which of these two approaches can better mitigate the hypothetical bias in hypothetical CEs? Finally, we also wish to know whether the priming task can make a difference in non-hypothetical choice experiments. In particular, we wish to test if either the honesty or neutral priming task can influence the WTP values from non-hypothetical choice experiments that are assumed to be true values.

To answer these four questions, we conducted two types of experiments (i.e., hypothetical and non-hypothetical choice experiments) with different treatments: hypothetical CE; non-hypothetical or real CE; hypothetical CE with cheap talk; hypothetical CE with neutral priming; hypothetical CE with honesty priming; non-hypothetical or real CE with neutral priming; and non-hypothetical or real CE with honesty priming.

[3] The cheap talk is also a mechanism with a priming component, but it contains explicit warnings, whereas with priming it is implicit.

Our study differs from previous studies in a number of ways. First, in contrast to Jacquement et al. (2011), who used the solemn oath, we used an implicit priming task to induce subjects to reveal their true preferences in CE. In the solemn oath approach, the person has to freely choose to "prime herself," while our honesty priming attempts to automatically activate a participant's honesty without the need for a direct consent. Second, in contrast to other studies, we conducted an artifactual field experiment using real products and a sample of participants drawn from a specific target population (i.e., consumers), instead of the standard subject pool of students. Finally, to the best of our knowledge this is the first study to test the use of an honesty priming technique in hypothetical choice experiments to mitigate hypothetical bias, as well as the effect of an honesty priming on the WTP values revealed from the real choice experiments.

The rest of the article is organized as follows: the next section discusses the experimental design and explains the rationale for including the various treatments. The following sections describe the econometric methodology and then the results. The final section discusses the importance and the implications of the findings.

General Design and Treatments' Description

We designed an artifactual field experiment using a non-standard subject pool (Harrison and List 2004). The target population was consumers instead of students, in an attempt to ensure that subjects were generally representative of shoppers in stores (Chang et al. 2009) and had experience with the good in question (List 2003). Moreover, to ensure that respondents had experience with the good, the target population was the primary food buyer in households that consumed the product of interest. The experiment was conducted in the region of Aragón (Spain), in the town of Zaragoza. The sample of participants was randomly selected from different locations across the city using a stratified sampling procedure by age, gender, and education level.

We implemented different treatments, and in accordance with Lusk and Schoeder (2004) we followed a between-subject approach where each respondent participates in only one of the treatments. The first treatment corresponded with the standard and commonly-used hypothetical choice experiment, while the second treatment is the non-hypothetical choice experiment where an economic incentive mechanism was added to induce respondents to truly reveal their preferences. These first two treatments represented the baseline treatments: the first is called the hypothetical baseline choice experiment (HB) and the second one is the real baseline choice experiment (RB). The third treatment consisted of a hypothetical CE with a cheap talk script that participants read before responding to the CE questions.[4] We refer to this as the cheap talk (CT) treatment. In the rest of the treatments, before participants responded to the choice tasks they were exposed to a subliminal priming technique called the "scrambled sentence test."[5] In this test, participants were asked to construct 24 grammatically-correct sentences out of a series of words presented in a scrambled order.[6] Two different priming exercises were defined, a neutral task and an honesty task.[7] In the honesty task the final sentences are related to honesty, fairness, and truthfulness (16 out of 24[8]), while in the neutral task, all the final sentences are not related to honesty concepts, but rather they correspond with general and basically known topics (e.g., the earth is round, summer is hot). We added the neutral priming to the honesty priming to test and ensure that the priming did not arise purely due to the nature of the scrambling task, but rather due to the activation of honesty concepts. Both priming exercises were introduced in both

[4] We used a generic, short and neutral cheap talk inspired by the one utilized by Cummings and Taylor (1999) and Silvia et al. (2011), which we modified and translated into Spanish (the English translation of our cheap talk is shown in the supplementary appendix on the OUP website).

[5] Psychologists use a conceptual priming technique to explore the influences of category representations. Conceptual priming is the activation of a cognitive representation in one context to unconsciously influence an unrelated context (Bargh 1996; Bargh and Chartrand 2000). An example of a non-conscious priming technique is the "scrambled test" (Bargh 1996).

[6] The particular sentences in the honesty priming and neutral priming were specified to fulfill three requirements: they should be easy to understand by participants with different educational backgrounds, be as short as possible to reduce the time for conducting the scrambled sentences, and be general and not politically sensitive to avoid impinging upon the sensitivities of some respondents. Before using these in the experiment, we validated the sentences using a pilot study of about 31 people. The objective of the pilot study was to examine if the participants had the skills to understand and conduct this test and to assess the time required to fill out the test.

[7] The honesty priming and neutral priming tasks can be seen in the supplementary appendix on the OUP website.

[8] We followed the Kay and Ross (2003) approach of using 24 sentences, where 16 are related to the concept representation.

the hypothetical and the real choice experiments.[9] Therefore, four additional treatments were conducted, the neutral priming in the hypothetical choice (HNP) and in the real choice (RNP), and the honesty priming in the hypothetical choice (HHP) and in the real choice experiment (RHP).

Regarding our four research questions, we had to build and test a series of hypotheses based on these treatments. To answer our first question we tested several hypotheses. The first hypothesis was defined as follows:

$$H0_1 : (WTP^{HHP} - WTP^{HB}) = 0, \text{ and}$$
$$H1_1 : (WTP^{HHP} - WTP^{HB}) < 0.$$

If $H0_1$ is rejected we might confirm that introducing honesty priming in the hypothetical CE reduces hypothetical bias because the WTP values in the honesty priming would be lower than in the standard baseline hypothetical CE.

However, before definitely confirming that introducing honesty priming in the hypothetical CE reduces the hypothetical bias, we need to ensure that other necessary conditions hold. First, we tested whether hypothetical bias indeed exists as follows:

$$H0_2 : (WTP^{RB} - WTP^{HB}) = 0$$
$$H1_2 : (WTP^{RB} - WTP^{HB}) < 0.$$

If $H0_2$ is rejected we can confirm that hypothetical bias indeed exists in hypothetical CEs.

Moreover, we have to ensure that the effect of the honesty priming task on the hypothetical WTP values did not arise purely due to the nature of the scrambling test, but rather due to the activation of honesty concepts. Hence, we expected that in the hypothetical CE, WTP values from the honesty priming and the neutral priming would be different (Bargh 1996; 2000; Bargh and Chartrand 20001; Kay and Ross 2003):

$$H0_3 : (WTP^{HHP} - WTP^{HNP}) = 0, \text{ and}$$
$$H1_3 : (WTP^{HHP} - WTP^{HNP}) < 0.$$

However, we expected the WTP values from the hypothetical baseline and the neutral priming to be the same:

$$H0_4 : (WTP^{HNP} - WTP^{HB}) = 0, \text{ and}$$
$$H1_4 : (WTP^{HNP} - WTP^{HB}) < 0.$$

Then $H0_3$ must be rejected, while $H0_4$ must not be rejected to ensure that the effect of the honesty priming on the WTP values is purely due to activating the honesty concept but not due to the scrambling test task.

To answer the second question, we checked whether hypothetical bias was totally reduced by honesty priming. Hence, we tested the following hypotheses:

$$H0_5 : (WTP^{RB} - WTP^{HHP}) = 0, \text{ and}$$
$$H1_5 : (WTP^{RB} - WTP^{HHP}) < 0.$$

If $H0_5$ is not rejected, then the WTP values from the real CE and the honesty priming applied to hypothetical CE are the same. In other words, using honesty priming in hypothetical CE drives the true WTP values.

To answer our third question, we tested the following hypothesis:

$$H0_6 : (WTP^{CT} - WTP^{HB}) = 0, \text{ and}$$
$$H1_6 : (WTP^{CT} - WTP^{HB}) < 0.$$

If $H0_6$ is rejected, we might confirm that introducing cheap talk in hypothetical CE reduces the hypothetical bias because the WTP values in cheap talk would be lower than in the standard baseline hypothetical CE. Moreover, we also tested whether cheap talk totally reduced the hypothetical bias as follows:

$$H0_7 : (WTP^{RB} - WTP^{CT}) = 0, \text{ and}$$
$$H1_7 : (WTP^{RB} - WTP^{CT}) < 0.$$

If $H0_7$ is rejected, this means that the hypothetical bias is not totally reduced by the use of the cheap talk in hypothetical CE.

Finally, to answer our last question, we tested whether the honesty priming influence of WTP values was revealed from non-hypothetical (incentivized) choice experiments. Hence, we

[9] To test whether subjects became aware of the manipulation, we asked subjects at the end of the experiment if they had noticed "a topic" from the words they were exposed to and the final sentences they had to write. Almost all participants did not notice the purpose of the honesty priming task.

tested the following hypotheses:

$$H0_8 : (WTP^{RHP} - WTP^{RB}) = 0, \text{ and}$$
$$H1_8 : (WTP^{RHP} - WTP^{RB}) < 0$$
$$H0_9 : (WTP^{RHP} - WTP^{RNP}) = 0, \text{ and}$$
$$H1_9 : (WTP^{RHP} - WTP^{RNP}) < 0$$
$$H0_{10} : (WTP^{RNP} - WTP^{RB}) = 0, \text{ and}$$
$$H0_{10} : (WTP^{RNP} - WTP^{RB}) < 0.$$

If we fail to reject $H0_8$, $H0_9$ and, $H0_{10}$, this would mean that the priming task would not make any difference in non-hypothetical (incentivized) choice experiments. This could then imply that when individuals are economically incentivized in choice experiments, they would reveal their "true" WTP values, and hence making them think more about honesty would not change their revealed WTP values.

Finally, as is standard practice in experiments of implicit priming manipulation, at the end of the experiment the subjects were asked if they noticed "a topic" from the words they were exposed to and the final sentences they had to write. All subjects (99%) reported unawareness of the goal-activation manipulation in either the neutral priming or the honesty priming treatments.

Experimental Procedure

The experiment consisted of two tasks plus the completion of a survey requesting basic information on socio-demographics. In the main task (choice experiment), participants faced different choice-set scenarios where they had to choose between two products with different attributes and prices, plus a no-buy option in the event they did not choose either of the two products for each scenario (Task I). Moreover, to be able to conduct an external validity analysis, we included an additional task (holdout task) (Task II) to determine whether the parameter estimates from the main task could predict actual preferences.[10] In the same vein as those utilized by Ding, Grewal, and Liechty (2005), our holdout task is a choice task but was held out of the utility estimation process since we used them to validate the model after the partworths are estimated.[11]

The experiment was conducted over several sessions with a maximum of 10-12 participants in each session. The sessions were conducted in a large room, and each participant was seated separately and far from other participants to avoid any communication among them during the experiment; all participants in the same session received the same treatment. In total, 383 subjects participated in all the treatments. The experiment was conducted as follows: first, participants were informed that they would receive €10 in cash at the end of the experiment for taking part. Then, each participant was assigned a unique ID number to guarantee his/her anonymity; before beginning the experiment they were asked to read and sign an informed consent form. In addition, respondents were asked to carefully study and inspect the different products in the choice sets, and a description of each of the products was presented and explained. The participants were then informed that they would take part in two sequential tasks (task I and task II), and that at the end of the experiment the monitor would randomly select one of the tasks to be binding. All the participants were informed that if task II was randomly selected as binding, they would buy the product they had chosen in this task and pay the corresponding price. Following Ding et al. (2005), we randomly selected the binding task and made task II non-hypothetical in all the treatments to compare the external predictive performance of the estimated partworths across the treatments.

In addition, participants in the non-hypothetical treatments were informed that if task I was randomly selected as binding, the experimenter would randomly select a number between 1 and 16 (total number of choice sets) to determine the binding choice set. The participants would then buy the product they had chosen in this binding choice set and pay the corresponding price, unless they picked the no-buy option. However, participants in the hypothetical treatments were informed that if task I was randomly selected, they would not have to buy any product.

Next, participants in the cheap talk treatment were asked to read the cheap talk scripts while participants in the neutral and

[10] Following Ding et al. (2005), participants in the holdout task faced eight different products, which were the remaining profiles from the original full fractional design that were not used in task I, plus a no-buy option. The holdout task was the same for all participants.

[11] We do not include the analysis of these data in the paper due to the small sample size and space constraints. These results are available from the authors upon request.

honesty treatments were asked to construct 24 grammatically correct sentences out of a series of scrambled words. The participants were then asked to start with task I and task II. Finally, once they finished both tasks and the numbers were drawn, participants were asked to complete a survey requesting basic information on socioeconomic and demographic characteristics.

Choice Set Design

The first step when implementing a choice experiment is to select the specific product to be analyzed. In accordance with Gracia, Loureiro, and Nayga (2011), we used a non-perishable product to isolate the effect of change in the food attributes from the organoleptic characteristics of the product (i.e., appearance and taste). We selected almonds as the product for our experiments because of its long tradition in the area where our experiment was conducted (the Aragón region of Spain), and because it is an important and frequently-consumed product for Spanish consumers. In particular, a package of 100 grams of untoasted almonds was selected.

The second step was to choose the attributes and levels to be used. One of the attributes was the price to allow for the calculation of the marginal WTPs. Four price levels were chosen to reflect those found in Spanish supermarkets (€1.35 , €1.84, €2.33, and €2.82 for a packet of 100 grams of untoasted almonds). A further aim of the experiment was to determine consumers' preferences for food products carrying two sustainability-related labels: organic and/or "food miles" labels. Therefore, our second attribute was the organic type of production with two levels: conventional (the product did not carry an organic label), and organic (the product carried the new EU organic label). The third attribute was the "food miles," representing the distance that the product had travelled (with four levels). The first level of travel had no label indicating the number of kilometers that the product travelled from the production facility. The second level corresponded with a package of almonds produced within 100 kilometers from Zaragoza city; in our case this meant the almonds were produced in the Zaragoza province. The third level denoted that the almonds were produced approximately 800 kilometers away from Zaragoza (i.e., the almonds were produced in some other Spanish region or neighboring country). The fourth level denoted that the almonds were produced approximately 2,000 kilometers from Zaragoza (i.e., produced outside of Spain but within Europe).

To avoid deceiving the participants, the almonds were either organic or conventional, and purchased from places matching the distance of transportation indicated in the "food miles" label. Table 1 shows the attributes and the levels used.

Since it is not realistic to force participants to choose one of the designed options (Louviere and Street 2000), each choice set included a no-buy option in addition to the two almond options. The choice set was designed following

Table 1. Attributes and Levels Used in the Choice Design

Attributes	Levels		
Price (€ per package)	1.35, 1.84, 2.33, and 2.82 (*PRICE*)		
EU organic label	No label EU organic label (*ORG*) 		
"Food miles" labels	No label	100 kilometers label (*km100*) 	800 kilometers label (*km800*) 	2,000 kilometers label (*km2000*) 

Street and Burgess (2007). To limit the number of choice sets and obtain an optimal design, we used an orthogonal main effect plan (OMEP) for developing the profiles in the first option (Street et al. 2005). We then added one of the generators suggested by Street and Burgess (2007) to obtain the profiles in the second option.[12] The orthogonal main effect plan was calculated using the SPSS orthoplan, which generated 16 profiles. We used these 16 profiles to obtain the products for the second option using one of the generators derived from the suggested difference vector (1 1 1) by Street and Burgess (2007) for 3 attributes with 4, 2, and 4 levels, respectively, and the two options. This design is 95.2% D-efficient compared to the optimal.

Econometric Methodology

To test our hypotheses we specified a utility function to calculate the marginal WTP based on Lancaster's consumer theory of utility maximization (Lancaster 1966), with consumer preferences for the attributes modeled within a random utility framework (McFadden 1974). Lancaster (1966) proposed that the total utility associated with the provision of a good can be decomposed into separate utilities for their component attributes. However, this utility is known to the individual but not to the researcher; the researcher observes some attributes of the alternatives but some components of the individual utility are unobservable and are treated as stochastic (Random Utility Theory). Thus, the utility is taken as a random variable where the utility from the n^{th} individual is based on the choice among j alternatives within the choice set in each of t choice occasions. In our empirical specification, the components of the utility function include the different attributes, as well as an alternative-specific constant (ASC) representing the no-buy option. The utility function is specified as follows:

$$U_{njt} = ASC + \alpha PRICE_{njt} + \beta_1 ORG + \beta_2 km100_{njt} + \beta_3 km800_{njt} + \beta_4 km2000_{njt} + e_{njt} \quad (1)$$

where n is the number of respondents, j denotes each of the three options available in the choice set, and t is the number of choice occasions. The ASC is a dummy variable indicating the selection of the no-buy option. It is expected that the constant ASC is negative and significant, indicating that consumers obtain lower utility from the no-buy option than for the designed alternatives. The price (*PRICE*) represents the price levels faced by consumers for the food product. Price is expected to have a negative impact on utility. As the organic attribute has two levels, one dummy variable was included, representing the organic label (*ORG*). In the same way, because the "food miles" attribute has four levels, three dummy variables were created (*km100, km800, and km2000*). Each of these variables takes the value +1 if the product carries the corresponding label, and 0 otherwise. Finally, ε_{njt} is an unobserved random term that is distributed following an extreme value type I (Gumbel) distribution, i.i.d. over alternatives, and independent of α and β, and the attributes that are known by the individual but unobserved and random from the researcher's perspective. Consumers are assumed to choose that alternative which provides the highest utility level from those available.

To estimate equation (1), we follow the state of the art and assume that price has a fixed coefficient and that the coefficients for the other attribute-level variables are random following a normal distribution. Following Layton and Brown (2000) and Revelt and Train (1998), we did not allow the price coefficient to vary in the population. Fixing the price coefficient ensures that the estimated willingness to pay will be normally distributed, and all respondents will have a negative price coefficient.

Instead of assuming homogenous preferences, we assumed that preferences are heterogeneous. We specified a Random Parameters Logit Model (RPL) considering a panel structure account for the fact that several choices were made by each individual (Train 2003). However, the assumption that the taste parameters are random but independently distributed may not be realistic. We can also expect that some attributes may be interdependent. To take these into account, the correlation structure of β_n was assumed to follow a multivariate normal distribution (normal with vector mean μ and variance-covariance matrix Ω) (Scarpa and Del Giudice 2004). We then estimated a Random Parameter Logit Model with correlated errors.

Because we are using various samples and treatments, it is important to investigate

[12] This design only allowed us to estimate the main effects.

whether differences in parameter estimates across samples are indeed due to the underlying preferences or to differences in variance. To accomplish this we follow Lusk and Schroeder (2004) to test if estimates from the RPL and the RPL with correlated errors are equivalent across pooled data from the hypothetical and non-hypothetical treatments using a test of the joint equality for the estimated parameters. The null hypothesis of the test is that the parameters are equal across the treatments. If this hypothesis is rejected, comparing the estimated WTP for each of the treatments would be appropriate because the error variance is constant within each sample, and it will be cancelled out when calculating the marginal WTPs. The test for equality is $-2(LL_j - \Sigma LL_i)$, which is distributed χ^2 with $K(M-1)$ degrees of freedom, where LL_j is the log likelihood value for the pooled data, LL_i are the log likelihood values for the different restricted models (treatments), K is the number of restrictions, and M is the number of treatments (Swait and Louviere 1993).

Based on the estimated coefficients from equation (1) we calculated the mean marginal WTP values for each attribute by taking the ratio of the mean parameter estimated for the non-monetary attributes to the mean price parameter, and multiplying by minus one. To test our hypotheses we used the combinatorial test suggested by Poe, Giraud, and Loomis (2005) to compare differences between estimated mean WTP in different treatments. This non-parametric test first required us to generate a distribution of 1,000 WTP estimates using, for example, the parametric bootstrapping method proposed by Krinsky and Robb (1986). The combinatorial test has also been applied by Lusk and Schroeder (2004), Lusk et al. (2003), Carlsson et al. (2005), Carlsson et al. (2007), and Tonsor and Shupp (2011).

In addition, to assess the robustness of our results to the econometric specification, we relaxed the assumption of fixed price coefficients and specified our utility in "WTP space" (Thiene and Scarpa 2009; Scarpa et al. 2008) instead of "preference space" (equation 1) to allow the price to be random.

Utility in the WTP space can be defined as follows (Greene 2012):

$$U_{njt} = \alpha(\theta_1 ASC + PRICE_{njt} + \theta_2 ORG + \theta_3 km100_{njt} + \theta_4 km800_{njt} + \theta_5 km2000_{njt}) + \varepsilon_{njt} \quad (2)$$

where $\theta_i = \beta_i/\alpha$ are already the willingness to pay estimates.

With this new utility specification, the estimates are directly considered the willingness to pay values, and we can test our hypotheses by pooling data for the two treatments involved in the particular hypothesis, and then specifying an extended utility with the appropriate set of treatment dummy variables, depending on the hypothesis to be tested:

$$\begin{aligned} U_{njt} = {} & \alpha(\theta_1 ASC + PRICE_{njt} + \theta_2 ORG \\ & + \theta_3 km100_{njt} + \theta_4 km800_{njt} \\ & + \theta_5 km2000_{njt}) + \delta_1(ORG_{njt} \\ & \times dtreat) + \delta_2(km100_{njt} \times dtreat) \\ & + \delta_3(km800_{njt} \times dtret) \\ & + \delta_4(km2000_{njt} \times dtreat) + \varepsilon_{njt} \end{aligned} \quad (3)$$

where *dtreat* is coded as 1 for the first treatment in the analyzed hypothesis, and 0 otherwise. We specified one extended utility function for each of the hypotheses to be tested. Thus, we estimated 10 extended utility functions using 10 different *dtreat* dummy variables.

The significance of the estimated δ and their signs will enable us to test differences in marginal WTP between the two treatments in the hypothesis to be analyzed. To accomplish this, we can use the t-test on the coefficient estimate.

Results

Table 2 reports the socio-demographic characteristics of the participants in the seven treatments. Because we require participants with similar socio-demographic characteristics across treatments to be able to compare results, we randomly recruited people by age, gender, and educational level for each of the treatments. We then used a chi-square test to determine if there were differences in socio-demographic profiles across treatments.

The results of these tests suggested that the null hypothesis of equality between the socio-demographic characteristics across treatment samples cannot be rejected at the 5% significance level for gender, age, education, and income. This result suggests that our randomization was successful in equalizing the characteristics of participants across the seven treatments.

To relax the homogeneity assumption of consumer preferences, we estimated equation (1)

Table 2. Sample Characteristics, Percentages

Variable definition	HB[a]	RB[b]	HHP[c]	HNP[d]	CT[e]	RHP[f]	RNP[g]
Gender							
Male	49.0	51.0	45.3	49.0	45.2	50	43.3
Female	51.0	49.0	54.7	51.0	54.7	50	56.7
Chi- Square (6) = 1.1269							
p-value = 0.98							
Age							
Between 18-35 years	24.5	28.3	30.2	26.4	26.4	25.9	25
Between 35-54 years	35.8	32.0	32.0	37.7	37.7	34.5	36.7
Between 55-64 years	16.9	18.8	15.0	15.8	15.0	17.2	15
More than 64 years	22.64	20.7	22.6	20.7	20.7	22.4	23.3
Chi- Square (18) = 1.6582							
p-value = 1.000							
Education of respondent							
Elementary School	26.4	24.5	24.5	22.6	22.6	25.9	23.3
High School	39.6	39.6	37.7	41.5	45.3	36.2	38.3
University	34.0	35.8	37.7	35.8	32.0	37.9	38.3
Chi- Square (12) = 1.5027							
p-value = 1.000							
Average household monthly net income							
Low Income (up to €1,500)	32.08	18.9	25.5	15.09	39.6	29.3	35
Medium Income (between €1,501-2,500)	28.30	47.2	30.2	33.9	26.4	34.5	23.3
High income (more than €2,500)	39.6	33.4	45.3	50.9	33.9	36.2	41.7
Chi- Square (12) = 18.3085							
p-value = 0.107							

Note: [a]HB denotes hypothetical baseline CE; [b]RB denotes non-hypothetical or real baseline CE; [c]HHP denotes hypothetical CE with honesty priming; [d]HNP denotes hypothetical CE with neutral priming; [e]CT denotes hypothetical CE with cheap talk; [f]RHP denotes non-hypothetical or real CE with honesty priming; [g]RNP denotes non-hypothetical or real CE with neutral priming.

using an RPL and an RPL with correlated errors where price is assumed to be fixed and the coefficients for the four attribute-level dummy variables are considered random following a normal distribution. Estimations were conducted using NLOGIT 5. To estimate these models, we used 100 Halton draws rather than pseudo-random draws since the former provides a more accurate simulation for the RPL model (Train 1999; Train 2003).

First, we tested the joint equality between the hypothetical and non-hypothetical estimates for both the RPL and the RPL with correlated errors using the likelihood ratio test. Table 3 reports the likelihood values for the pooled and segmented samples (treatments), together with the tests of equality for the RPL and the RPL with correlated errors. The results indicate that the joint null hypotheses of equality between the hypothetical and non-hypothetical treatments (LR = 78.61 for the RPL and LR = 77.4 for the RPL with correlated matrix) are rejected, suggesting that it would be appropriate to compare the estimated WTPs between hypothetical and non-hypothetical treatments. Moreover, the hypothesis of equality across hypothetical choice treatments is also rejected (LR = 86.26 for the RPL and LR = 70.06 for the RPL with correlated errors), as is the hypothesis of equality across non-hypothetical choice treatments (LR = 41.63 for the RPL and LR = 46.10 for the RPL with correlated errors). We can then conclude that comparing the estimated parameters from the various treatments is appropriate when estimating the models separately.[13]

Finally, if we examine the log likelihood values, we see that the best values are found in the RPL model with correlated errors across the different treatments. Hence, the best fit for our data seems to be the RPL model with correlated errors, and hence we used this model to calculate the WTPs for the different treatments to test our research hypotheses.

[13] To conserve space, the estimated parameters from the models are reported in the supplementary appendix on the OUP website. Note that all mean coefficients and the corresponding WTPs are statistically significant, except for some coefficients of the KM800 variable.

Table 3. Hypothesis Tests of Equality across Hypothetical and Non-Hypothetical Treatments

Hypothesis Tests of Equality	N. obs[a]	Random Parameter Logit Model		Random Parameter Logit Model with correlated errors	
		Log Likelihood	*p-value*	Log Likelihood	*p-value*
All treatments	18,384	−4, 931.29		−4, 808.71	
All hypothetical treatments	10,176	−2, 591.98		−2, 530.33	
All non-hypothetical treatments	8,208	−2, 300.81		−2, 239.70	
H0 = Test of equality between hypothetical and non-hypothetical treatments		78.61	0.000	77.4	0.004
All hypothetical treatments	10,176	−2, 591.16		−2530.33	
Hypothetical baseline CE (HB)	2,544	−632.98		−621.92	
Hypothetical CE with honesty priming (HHP)	2,544	−661.92		−654.39	
Hypothetical CE with neutral priming (HNP)	2,544	−626.38		−611.4	
Hypothetical CE with cheap talk (CT)	2,544	−626.72		−607.59	
H0 = Test of equality across hypothetical treatments		86.26	0.000	70.06	0.020
All non-hypothetical treatments	8,208	2, 300.81		−2, 239.70	
Non-hypothetical CE baseline (RB)	2,544	−697.83		−666.49	
Non-hypothetical CE with honesty priming (RHP)	2,784	−775.04		−758.07	
Non-hypothetical CE with neutral priming (RNP)	2,880	−807.12		−792.05	
H0 = Test of equality across non hypothetical treatments		41.63	0.030	46.10	0.050

Note: [a] indicates number of observations.

However, since the parameters in the RPL models were normalized by the unobserved portion of utility, mean parameter estimates cannot be directly compared across treatments (Revelt and Train 1998).

Tables 4 and 5 report the marginal WTPs across the seven treatments and the corresponding hypothesis tests using the nonparametric combinational method of Poe, Giraud and Loomis (2005), with 1,000 Krinsky-Robb (1986) bootstrapped WTP estimates. To test our 10 hypotheses, we used one-sided tests because of the type of alternative hypotheses we considered.

First, table 4 shows that our first hypothesis ($H0_1$: $(WTP^{HHP} - WTP^{HB}) = 0$; $H1_1$: $(WTP^{HHP} - WTP^{HB}) < 0$) is rejected in three of the four analyzed labels, indicating that marginal WTPs from the CE using the honesty priming task are lower than those from our baseline hypothetical CE. This result implies that the honesty priming task could reduce the hypothetical bias in hypothetical choice experiments. However, to definitively answer our first research question, we have to be sure that the necessary hypotheses, such as $H0_2$, $H0_3$, and $H0_5$ must also be rejected, while $H0_4$ must not be rejected. Our analyses show that the second hypothesis ($H0_2$: $(WTP^{RB} - WTP^{HB}) = 0$; $H1_2$: $(WTP^{RB} - WTP^{HB}) < 0$) is rejected in the four analyzed labels. This result confirms that hypothetical bias exists in hypothetical CE because WTPs in the hypothetical CE are greater than the WTPs in non-hypothetical CE. In particular, the marginal WTPs in table 4 indicate that the participants overstated their WTPs across the labels by an average factor of about 1.40. This result is similar to those obtained by Murphy et al. (2005) and Lusk and Schroeder (2004), who found a factor of around 1.20. Our third hypothesis, ($H0_3$: $(WTP^{HHP} - WTP^{HNP}) = 0$; $H1_3$: $(WTP^{HHP} - WTP^{HNP}) < 0$), is rejected in two of the four analyzed labels, indicating that the WTP values from the honesty priming treatment are less than or equal to those from the neutral priming. In contrast, we did not reject the fourth hypothesis ($H0_4$: $(WTP^{HNP} - WTP^{HB}) = 0$; $H0_4$: $(WTP^{HNP} - WTP^{HB}) < 0$), which confirms that priming

Table 4. Marginal WTP Values (€/100 grams) across Treatments and Hypotheses Tests $H0_1 - H0_4$

Hypotheses Tests	*ORG*	*km100*	*km800*	*km2000*
$H0_1$[a] $(WTP^{HHP} - WTP^{HB}) = 0$				
[c]WTP^{HHP}	0.46	0.72	0.01	−0.57
[d]WTP^{HB}	0.87	1.11	0.27	−0.52
p-value[b]	0.00	0.03	0.10	0.42
$H0_2$[a] $(WTP^{RB} - WTP^{HB}) = 0$				
[e]WTP^{RB}	0.61	0.74	−0.18	−1.19
[d]WTP^{HB}	0.87	1.11	0.27	−0.52
p-value[b]	0.09	0.06	0.04	0.02
$H0_3$[a] $(WTP^{HHP} - WTP^{HNP}) = 0$				
[c]WTP^{HHP}	0.46	0.72	0.01	−0.57
[f]WTP^{HNP}	0.60	1.06	0.29	−0.65
p-value[b]	0.18	0.04	0.07	0.38
$H0_4$[a] $(WTP^{HNP} - WTP^{HB}) = 0$				
[f]WTP^{HNP}	0.60	1.06	0.29	−0.65
[d]WTP^{HB}	0.87	1.11	0.27	−0.52
p-value[b]	0.09	0.43	0.46	0.33

Note: [a]$H0_1, H0_2, H0_3$, and $H0_4$ represent our first four hypotheses for the first research question mentioned in the introduction section. [b]*p*-values were estimated using the combinational method of Poe, Giraud, and Loomis (2005) with 1,000 Krinsky-Robb (1986) bootstrapped WTP estimates. The *p*-value reports results of the one-sided test for our hypotheses for each corresponding pair of attributes; [c]WTP^{HHP} indicates bootstrapped WTP estimates in hypothetical CE with honesty priming; [d]WTP^{HB} denotes bootstrapped WTP estimates in hypothetical baseline CE; [e]WTP^{RB} denotes bootstrapped WTP estimates in non-hypothetical CE; [f]WTP^{HNP} denotes bootstrapped WTP estimates in hypothetical CE with neutral priming.

Table 5. Marginal WTP Values (€/100 grams) across Treatments and Hypotheses Tests $H0_5 - H0_7$

Hypotheses Tests	*ORG*	*km100*	*km800*	*km2000*
$H0_5$[a] $(WTP^{RB} - WTP^{HHP}) = 0$				
[d]WTP^{RB}	0.61	0.74	−0.18	−1.19
[c]WTP^{HHP}	0.46	0.72	0.01	−0.57
p-value[b]	0.14	0.45	0.20	0.02
$H0_6$[a] $(WTP^{CT} - WTP^{HB}) = 0$				
[f]WTP^{CT}	0.57	0.93	0.38	−0.39
[d]WTP^{HB}	0.87	1.11	0.27	−0.52
p-value [b]	0.06	0.21	0.31	0.33
$H0_7$[a] $(WTP^{RB} - WTP^{CT}) = 0$				
[e]WTP^{RB}	0.61	0.74	−0.18	−1.19
[f]WTP^{CT}	0.57	0.93	0.38	−0.39
p-value [b]	0.40	0.17	0.01	0.00

Note: [a]$H0_5$; $H0_6$, and $H0_7$ represent our second set of hypotheses for the second and third research questions mentioned in the introduction section; [b]*p*-values were estimated using the combinational method of Poe, Giraud, and Loomis (2005) with 1,000 Krinsky-Robb (1986) bootstrapped WTP estimates; *p*-value reports results of the one-sided test for our hypotheses for each corresponding pair of attributes; [c]WTP^{HHP} indicates bootstrapped WTP estimates in hypothetical CE with honesty priming; [d]WTP^{HB} indicates bootstrapped WTP estimates in hypothetical baseline CE; [e]WTP^{RB} means bootstrapped WTP estimates in non-hypothetical CE; [f]WTP^{CT} indicates bootstrapped WTP estimates in hypothetical CE with cheap talk.

Table 6. Marginal WTP Values (€/100 grams) across Treatments and Hypotheses Tests $H0_8$–$H0_{10}$

Hypotheses Tests	*ORG*	*km100*	*km800*	*km2000*
$H0_8$ [a] $(WTP^{RHP} - WTP^{RB}) = 0$				
[c]WTP^{RHP}	0.40	0.55	−0.30	−0.93
[e]WTP^{RB}	0.61	0.74	−0.18	−1.19
p-value [b]	0.08	0.15	0.33	0.22
$H0_9$ [a] $(WTP^{RHP} - WTP^{RNP}) = 0$				
[c]WTP^{RHP}	0.40	0.55	−0.30	−0.93
[d]WTP^{RNP}	0.61	0.55	−0.22	−0.74
p-value [b]	0.06	0.49	0.38	0.27
$H0_{10}$ [a] $(WTP^{RNP} - WTP^{RB}) = 0$				
[d]WTP^{RNP}	0.61	0.55	−0.22	−0.74
[e]WTP^{RB}	0.61	0.74	−0.18	−1.19
p-value [b]	0.48	0.13	0.44	0.07

Note: [a]$H0_8$; $H0_9$ and, $H0_{10}$ represent our last three hypotheses for the fourth research question mentioned in the introduction section; [b]*p*-values were estimated using the combinational method of Poe, Giraud, and Loomis (2005) with 1,000 Krinsky-Robb (1986) bootstrapped WTP estimates; *p*-value reports results of the one-sided test for our hypotheses for each corresponding pair of attributes; [c]WTP^{RHP} indicates bootstrapped WTP estimates in non- hypothetical CE with honesty priming; [d]WTP^{RNP} indicates bootstrapped WTP estimates in non- hypothetical CE with neutral priming; [e]WTP^{RB} indicates bootstrapped WTP estimates in non-hypothetical CE.

effects do not arise purely due to the nature of the scrambling task, but rather due to the activation of honesty concepts, given that the WTP estimates in the hypothetical neutral priming treatment (HNP) are not statistically different from WTPs in the hypothetical baseline treatment (HB). In other words, the HNP did not induce either a task or priming effect. It also suggests that the scrambled sentence task in itself did not influence the participants' subsequent choice tasks. So in general, the results of these hypotheses' tests generally indicate that hypothetical bias exists in hypothetical choice experiments and that the honesty priming task can mitigate this bias.

To answer our second research question regarding how much the hypothetical bias is reduced by the use of honesty priming in hypothetical CE, we can notice that the fifth hypothesis ($H0_5$: $(WTP^{RB} - WTP^{HHP}) = 0$; $H1_5$: $(WTP^{RB} - WTP^{HHP}) < 0$) is not rejected in three of the four labels (table 5), suggesting that the WTPs from the real CE and honesty priming in hypothetical CEs are generally the same.

Now we attempt to answer our third question on whether the cheap talk script could mitigate the hypothetical bias in hypothetical CE, and to what extent. We notice in table 5 that the sixth hypothesis ($H0_6$: $(WTP^{CT} - WTP^{HB}) = 0$; $H1_6$: $(WTP^{CT} - WTP^{HB}) > 0$) is rejected only in one out of the four labels, indicating that in most cases, the cheap talk approach was not able to mitigate the hypothetical bias in hypothetical CEs. Moreover, the seventh hypothesis ($H0_7$: $(WTP^{RB} - WTP^{CT}) = 0$; $H1_7$: $(WTP^{RB} - WTP^{CT}) < 0$) is rejected in two out of the four labels, suggesting that the WTP values from the cheap talk treatment are higher than or equal to the WTP values from the real CE.

Finally, we are also able to test our fourth research question on whether a priming task would make any difference in non-hypothetical (incentivised) choice experiments (table 6). Results indicate that the eight ($H0_8$: $(WTP^{RHP} - WTP^{RB}) = 0$; $H1_8$: $(WTP^{RHP} - WTP^{RB}) < 0$), ninth ($H0_9$: $WTP^{RHP} - WTP^{RNP}) = 0$; ($H0_9$:$WTP^{RHP} - WTP^{RNP}) < 0$), and tenth ($H0_{10}$:$WTP^{RNP} - WTP^{RB}) = 0$; $H1_{10}$: $WTP^{RNP} - WTP^{RB}) < 0$) hypotheses are not rejected in three out the four labels, confirming that the WTP values from non-hypothetical CE are not affected by either the neutral or honesty priming task.

Finally, as mentioned above, we assessed the robustness of our hypotheses' test results for the econometric specification by relaxing the assumption of fixed price coefficients and specifying our utility in the WTP space. Tables 7, 8

Table 7. Robustness Tests in WTP Space (€/100 grams): Hypotheses Tests $H0_1 - H0_4$

Hypotheses Tests	*Coefficient*[b]	*Standard error*	*p-value*
$H0_1$[a]			
$(WTP^{HHP} - WTP^{HB}) = 0$			
$ORG \times dtreat_{HHP}$	−0.681***	0.375	0.006
$km100 \times dtreat_{HHP}$	−0.193	0.329	0.558
$km800 \times dtreat_{HHP}$	−0.801**	0.383	0.036
$km2000 \times dtreat_{HHP}$	−1.327***	0.457	0.004
$H0_2$[a]			
$(WTP^{RB} - WTP^{HB}) = 0$			
$ORG \times dtreat_{RB}$	−0.703**	0.369	0.057
$km100 \times dtreat_{RB}$	−0.950**	0.331	0.041
$km800 \times dtreat_{RB}$	−1.181**	0.482	0.014
$km2000 \times dtreat_{RB}$	−1.460***	0.613	0.000
$H0_3$[a]			
$(WTP^{HHP} - WTP^{HNP}) = 0$			
$ORG \times dtreat_{HHP}$	−0.506**	0.178	0.005
$km100 \times dtreat_{HHP}$	−0.653*	0.382	0.087
$km800 \times dtreat_{HHP}$	−0.613	0.382	0.119
$km2000 \times dtreat_{HHP}$	0.260	0.601	0.664
$H0_4$[a]			
$(WTP^{HNP} - WTP^{HB}) = 0$			
$ORG \times dtreat_{HNP}$	−0.235	0.316	0.457
$km100 \times dtreat_{HNP}$	−0.076	0.336	0.819
$km800 \times dtreat_{HNP}$	−0.062	0.379	0.869
$km2000 \times dtreat_{HNP}$	−0.466	0.4881	0.338

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively;
[a]$H0_1, H0_2, H0_3$, and $H0_4$ represent our first four hypotheses for the first question mentioned in the introduction section; [b]designates the effects of the treatment (*dtreat*) on the marginal WTP estimate.

Table 8. Robustness Tests in WTP Space (€/100 grams): Hypotheses Tests $H0_5 - H0_7$

Hypotheses Tests	*Coefficient*[b]	*Standard error*	*p-value*
$H0_5$[a]			
$(WTP^{RB} - WTP^{HHP}) = 0$			
$ORG \times dtreat_{RB}$	−0.024	0.271	0.928
$km100 \times dtreat_{RB}$	−0.510*	0.295	0.084
$km800 \times dtreat_{RB}$	−0.414	0.296	0.167
$km2000 \times dtreat_{RB}$	−0.955***	0.484	0.048
$H0_6$[a]			
$(WTP^{CT} - WTP^{HB}) = 0$			
$ORG \times dtreat_{CT}$	−0.419	0343	0.221
$km100 \times dtreat_{CT}$	−0.277	0.327	0.398
$km800 \times dtreat_{CT}$	−0.314	0.393	0.423
$km2000 \times dtreat_{CT}$	−0.621	0.549	0.257
$H0_7$[a]			
$(WTP^{RB} - WTP^{CT}) = 0$			
$ORG \times dtreat_{RB}$	−0.258	0.281	0.359
$km100 \times dtreat_{RB}$	−0.579**	0.291	0.047
$km800 \times dtreat_{RB}$	−0.957**	0.457	0.036
$km2000 \times dtreat_{RB}$	−0.970***	0.492	0.000

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively; [a]$H0_5, H0_6$, and $H0_7$ represent our second-set hypotheses for the second and third research questions mentioned in the introduction section; [b]indicates the effects of the treatment (*dtreat*) on the marginal WTP estimate.

Table 9. Robustness Tests in WTP Space (€/100 grams): Hypotheses Tests $H0_8 - H0_{10}$

Hypotheses Tests	*Coefficient*[b]	*Standard error*	*p-value*
$H0_8$[a]			
$(WTP^{RHP} - WTP^{RB}) = 0$			
$ORG \times dtreat_{RHP}$	0.017	0.248	0.991
$km100 \times dtreat_{RHP}$	0.070	0.354	0.842
$km800 \times dtreat_{RHP}$	0.135	0.395	0.972
$km2000 \times dtreat_{RHP}$	0.284	0.403	0.482
$H0_9$[a]			
$(WTP^{RHP} - WTP^{RNP}) = 0$			
$ORG \times dtreat_{RHP}$	−0.281	0.223	0.204
$km100 \times dtreat_{RHP}$	0.134	0.267	0.617
$km800 \times dtreat_{RHP}$	−0.190	0.372	0.609
$km2000 \times dtreat_{RHP}$	−0.328	0.335	0.326
$H0_{10}$[a]			
$(WTP^{RNP} - WTP^{RB}) = 0$			
$ORG \times dtreat_{RNP}$	0.255	0.261	0.329
$km100 \times dtreat_{RNP}$	−0.060	0.292	0.835
$km800 \times dtreat_{RNP}$	0.119	0.419	0.775
$km2000 \times dtreat_{RNP}$	0.567	0.428	0185

Note: ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively; [a]$H0_8$; $H0_9$, and $H0_{10}$ represent our last set of hypotheses for the fourth question mentioned in the introduction section; indicates the effects of the treatment (*dtreat*) on the marginal WTP estimate.

and 9 exhibit the estimated parameters and the corresponding t-values for the dummy treatment variables (δ) needed to test our hypotheses. Because we use the second WTP values in the null hypothesis as the reference levels, we expected that all the estimated coefficients will be negative in accordance with the alternative set of hypotheses. Our results confirmed these expectations because all the estimated coefficients are negative. We can notice that results of the hypotheses' tests are similar to the results obtained above using the Poe, Giraud, and Loomis (2005) test. In particular, hypothesis one is also rejected in three of the four analyzed labels; hypothesis two is rejected in all four analyzed labels; and hypothesis three is rejected in two of the four labels. However, we did not reject hypothesis four in all four labels, which is also quite consistent with and earlier Poe, Giraud, and Loomis (2005) test where this hypothesis was not rejected for three labels. In addition, hypothesis five is not rejected in two of the four labels, and hypothesis six is not rejected in all four labels as before using the Poe, Giraud, and Loomis (2005) test. Hypothesis seven is rejected in three labels compared to two labels using the Poe, Giraud, and Loomis (2005) test.

Finally, hypotheses 8, 9, and 10 are not rejected in all four labels compared to three labels using the Poe, Giraud, and Loomis (2005) test.

Conclusions

Undoubtedly, the choice experiment (CE) approach is the most widely used stated preference method for valuing products or attributes in the applied economics and marketing literature. However, a major issue that has challenged researchers who use this method is the hypothetical bias issue. Due to overwhelming evidence indicating the existence of hypothetical bias in stated valuation research, non-hypothetical experimental valuation methods have surfaced in the literature, including non-hypothetical choice experiments (see Gracia, Loureiro, and Nayga 2011). However, the problem with using non-hypothetical CE is that one actually needs to have all the product profiles in the choice sets produced and ready to be exchanged for money to align the mechanism incentive. While making the CE non-hypothetical is noteworthy, it is not always feasible to adopt this method given the challenges of producing all the product profiles being tested. In addition to being a relatively new method, this is probably the reason why the percentage of CE studies conducted non-hypothetically is significantly smaller than the percentage of CE studies done hypothetically.

Due to the challenge of using the non-hypothetical version of CE, a number of studies have tested the effectiveness of ex-ante calibration methods, for example the cheap

talk script for reducing hypothetical bias in CE studies, with mixed results. One of the possible reasons from a social psychology perspective is that cheap talk is a necessary but not a sufficient condition to trigger proper behavior. Therefore, some studies have proposed alternative ex-ante calibration mechanisms to elicit honest and truthful answers using a commitment devise such as a solemn oath (Jaquement et al. 2011; Jaquement et al. 2013).

Our approach was constructed and implemented in the same spirit as the solemn oath studies. We proposed and tested an *ex ante* instrument from the social psychology field that has not been tried before in CE studies: the honesty priming task. In particular, we tried to address four questions. The first question is whether exposure to honesty concepts could unconsciously activate honesty among subjects and let them respond more truthfully, and in turn mitigate potential hypothetical bias in hypothetical choice experiments. Our results generally suggest that the honesty priming task can indeed reduce hypothetical bias in hypothetical choice experiments. Specifically, we found that marginal WTPs in the honesty priming treatment are significantly lower on average than those in our baseline hypothetical CE treatment. Given that our results from the baseline hypothetical CE are not significantly different from those in the hypothetical CE with neutral priming, this finding implies that the change in behavior in the honesty priming treatment is due only to the honesty priming task, and not to the nature of the scrambling sentence test.

Since we found a positive answer to the first question, our next question was how much honesty priming can reduce hypothetical bias. Specifically, we tested if the marginal WTPs from the honesty priming hypothetical choice experiment are comparable to the marginal WTPs from the non-hypothetical choice experiment. Our results generally suggest that values from hypothetical CE with honesty priming are not significantly different from those from non-hypothetical CE. We were also able to answer our third question, which related to the ability of a cheap talk script to reduce the hypothetical bias in hypothetical CE. Our results suggest that the cheap talk script was not able to mitigate the hypothetical bias in hypothetical CE, and that WTP values from hypothetical CE with cheap talk are generally not as close to the real CE as the WTP values from the honesty priming applied to hypothetical CE.

These findings seem to suggest that untruthful choice revelation is one of the major reasons for the occurrence of hypothetical bias in hypothetical CE studies, given the effectiveness of the honesty priming task. Admittedly, this does not necessarily mean that the honesty priming task in itself could not trigger some other psychological effect that could address the other reasons for the existence of hypothetical bias (e.g., some subjects may not exactly know their WTP values), but the results generally indicate that untruthful revelation is a major source of the bias.

Finally, we also addressed our fourth research question on whether a priming task would make any difference in non-hypothetical (incentivized) choice experiments. We did not find statistically significant differences between WTP values from any of the non-hypothetical CE treatments, with or without the priming tasks.

Our findings hold some promise for using honesty priming to mitigate hypothetical bias in hypothetical choice experiments. This is an important finding considering that it is not always possible to conduct a choice experiment non-hypothetically as discussed above. Our finding implies that if it is not feasible to conduct a choice experiment non-hypothetically, then one could potentially consider the use of honesty priming to help mitigate potential hypothetical bias in hypothetical choice experiment studies. Moreover, we also found that WTP values from the real CE are not affected by either honesty priming or neutral priming, suggesting that WTP values revealed from the real CE could indeed be the true WTP values as assumed. However, our work represents only one study and therefore must be replicated in other settings or contexts to test the robustness of our findings.

References

Alfnes, F., A.G. Guttormsen, G. Steine, and K. Kolstad. 2006. Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives. *American Journal of Agricultural Economics* 88 (4): 1050–1061.

Bargh, J. 1990. Auto-Motives: Preconscious Determinants of Thought and Behaviour. In *Handbook of Motivation and Cognition*, ed. E.T. Higgins and R.M. Sorrentino, 93–130. New York: Guilford Press.

———. 1996. Automaticity in Social Psychology. In *Social Psychology: Handbook of Basic Principles*, ed. E.T. Higgins and A.W. Kruglanski, 169–183. New York: Guilford Press.

———. 1997. The Automaticity of Everyday Life. In *Advances in Experimental Social Psychology*, ed. R.S. Wyer, Jr., 1–61. Mahwah, NJ: Lawrence Erlbaum Associates.

Bargh, J., P.M. Gollwitzer, A.Y. Lee-Chai, K. Barndollar, and R. Troetschel. 2001. The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals. *Journal of Personality and Social Psychology* 81: 1014–1027.

Benjamin, D.J., J.J. Choi, and A.J. Strickland. 2010. Social Identity and Preferences. *American Economic Review* 100: 1913–1928.

Benjamin, D.J., J.J. Choi, and G. Fisher. 2011. Religious Identity and Economic Behavior. Working paper, The National Bureau of Economic Research, Cambridge, MA.

Brummett, R.G., R. Jr. Nayga, and X. Wu. 2007. On the Use of Cheap Talk in New Product Valuation. *Economics Bulletin* 2: 1–9.

Cameron, T., G. Poe, R. Ethier, and W. Schulze. 2002. Alternative Non-Market Value-Elicitation Methods: Are the Underlying Preferences the Same? *Journal of Environmental Economics and Management* 44: 391–425.

Carlsson, F., and P. Martinsson. 2001. Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to the Valuation of the Environment. *Journal of Environmental and Economics Management* 41 (2): 179–192.

Carlsson, F., P. Frykblom, and C.J. Lagerkvist. 2005. Using Cheap Talk as a Test of Validity in Choice Experiments. *Economics Letters* 89: 147–152.

———. 2007. Consumer Willingness to Pay for Farm Animal Welfare: Mobile Abattoirs versus Transportation to Slaughter. *European Review of Agricultural Economics* 34 (3): 321–344.

Chang, J.B., J. Lusk, and F.B. Norwood. 2009. How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behaviour? *American Journal of Agricultural Economics* 91 (2): 518–534.

Chartrand, T.L., J. Huber, B. Shiv, and R.J. Tanner. 2008. Nonconscious Goals and Consumer Choice. *Journal of Consumer Research* 35: 189–201.

Cummings, R.G., and L.O. Taylor. 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *The American Economic Review* 89: 649–665.

Dee, T.S. 2010. Stereotype Threat and the Student-Athlete. Working paper, The National Bureau of Economic Research, Cambridge, MA.

Ding, M., R. Grewal, and J. Liechty. 2005. Incentive-Aligned Conjoint Analysis. *Journal of Marketing Research* 42: 67–82.

Drouvelis, M., R. Metcalfe, and N. Powdthavee. 2010. Priming Cooperation in Social Dilemma Games. Working paper, The Institute for the Study of Labor (IZA), Bonn, Germany.

Gracia, A., M.L. Loureiro, and R. Nayga, Jr. 2011. Are Valuation from Non-Hypothetical Choice Experiments different from those of Experimental Auctions? *American Journal of Agricultural Economics* 93 (5): 1358–133.

Greene, W.H. 2011. *NLOGIT Version 5 Reference Guide*. United States: Econometric Software, Inc.

Harrison, G.W. 2006. Experimental Evidence on Alternative Environmental Valuation Methods. *Environmental and Resource Economics* 34: 125–162.

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

Jacquemet, N., R.V. Joule, S. Luchini, and J.F. Shogren. 2013. Preference Elicitation Under Oath. *Journal of Environmental Economics and Management* 65 (1): 110–132.

Jacquemet, N., A.G. Jame, S. Luchini, and J.F. Shogren. 2011. Social Psychology and Environmental Economics: A New Look at Ex-Ante Corrections of Biases Preference Evaluation. *Environmental Resource Economics* 48: 413–433.

~~Johansson-Stenman, O., H. Svedsäter. 2008. Measuring Hypothetical Bias in Choice Experiments: The Importance of Cognitive Consistency. *The Berkeley Electronic Journal of Economic Analysis and Policy* 8 (1): 1935–1682.~~

Joule, R., F. Bernard, and S. Halimi-Falkowicz. 2008. Promoting Ecocitenship: in Favour of Binding Communication. *International Scientific Journal for Alternative Energy and Ecology* 6 (62): 214–218.

Kay, A.C., and L. Ross. 2003. The Perceptual Push: The Interplay of Implicit Cues and Explicit Situational Construals on Behavioral Intentions in the Prisoners' Dilemma. *Journal of Experimental Social Psychology* 39: 634–643.

Krinsky, I., and A.L. Robb. 1986. On Approximating the Statistical Properties of Elasticities. *Review of Economics and Statistics* 64: 715–719.

Lancaster, K.A. 1966. New Approach to Consumer Theory. *Journal of Political Economy* 74: 132–157.

Layton, D.F., and G. Brown. Heterogeneous Preferences Regarding Global Climate Change. *The Review of Economics and Statistics* 82: 616–624.

List, J.A., and G.A. Gallet. 2001. What Experimental Protocol Influence Disparities Between Actual and Hypothetical State Value? *Environmental and Resource Economics* 20: 241–254.

List, J.A. 2001. Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards. *American Economic Review* 91: 1498–1507.

List, J.A. 2003. Does Market Experience Eliminate Market Anomalies? *The Quarterly Journal of Economics* 118: 41–71.

Loomis, J., P. Bell, H. Cooney, and C. Asmus. 2009. A Comparison of Actual and Hypothetical Willingness to Pay of Parents and Non-Parents for Protecting Infant Health: The Case of Nitrates in Drinking Water. *Journal of Agricultural and Applied Economics* 41 (3): 697–712.

Loomis, J. 2011. What's to Know About Hypothetical Bias in Stated Preferences Valuation Studies? *Journal of Economic Survey* 25 (2): 363–370.

Louviere, J.J., and D. Street. 2000. Stated-Preference Methods. In *Handbook of Transport I: Transport Modelling*, ed. D.A. Hensher and K.J. Button. Amsterdam: Pergamon Press.

Lusk, J.L., J. Roosen, and J.A. Fox. 2003. Demand for Beef from Cattle Administered Growth Hormones or Fed Genetically Modified Corn: A Comparison of Consumers in France, Germany, the United Kingdom, and the United States. *American Journal of Agricultural Economics* 85: 16–29.

Lusk, J.L., and T.C. Schroeder. 2004. Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. *American Journal of Agricultural Economics* 86: 567–582.

Lusk, J.L., D. Fields, and W. Prevett. 2008. An Incentive Compatible Conjoint Ranking Mechanism. *American Journal of Agricultural Economics* 90 (2): 487–498.

Maxwell, S., P. Nye, and N. Maxwell. 1999. Less Pain, Same Gain: The Effects of Priming Fairness in Price Negotiations. *Psychology & Marketing* 16: 545–562.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press.

McKay, R., C. Efferson, H. Whitehouse, and E. Fehr. 2011. Wrath of God: Religious Primes and Punishment. *Proceedings of the Royal Society B – Biological Sciences* 278: 1858–1863.

Murphy, J.J., J.A. Geoffrey, T.H. Stevens, and D. Weatherhead. 2005. Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Journal of Environmental and Resource Economics* 30: 313–325.

Poe, G.L., K.L. Giraud, and J.B. Loomis. 2005. Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87 (2): 353–365.

Revelt, D., and K. Train. 1998. Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level. *The Review of Economics and Statistics* 80: 647–57.

Scarpa, R., and T. Del Giudice. 2004. Market Segmentation Via Mixed Logit: Extra-Virgin Olive Oil in Urban Italy. *Journal of Agricultural and Food Industrial Organization* 2 (~~7~~): 7.

Scarpa, R., M. Thiene, and K. Train. 2008. Utility Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps. *American Journal of Agricultural Economics* 90 (4): 994–1010.

Silva, A., R. Jr. Nayga, B.L. Campbell, and L.J. Park. 2011. Revisiting Cheap Talk with New Evidence from a Field Experiment. *Journal of Agricultural and Resource Economics* 36 (2): 280–291.

Smith, V.L. 1976. Experimental Economics: Induced Value Theory. *American Economic Review* 66 (~~2~~): 274–79.

Street, D., L. Burgess, and J.J. Louviere. 2005. Quick and Easy Sets: Constructing Optimal and Nearly Optimal Stated

Choice Experiment. *International Journal of Research in Marketing* 22: 459–470.

Street, D., and L. Burgess. 2007. *The Construction of Optimal Stated Choice Experiments*. New Jersey: John Wiley & Sons Inc.

Swait, J., and J.J. Louviere. 1993. The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models. *Journal of Marketing Research* 30: 305–314.

Tiene, M., and R. Scarpa. 2009. Deriving and Testing Efficient Estimates of WTP Distributions in Destination Models. *Environmental Resource Economics* 44: 379–395.

Tonson, G.T., and R.S. Shupp. 2011. Cheap Talk Scripts Online Choice Experiment: Looking beyond the Mean. *American Journal of Agricultural Economics* 93 (4): 1015–1031.

Train, K. 1999. Halton sequences for Mixed Logit. Working Paper, University of California, Berkley.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

~~Volinskiy, D., W. Adamowicz, M. Veeman, and L. Srivastava. 2009. Does Choice Context Affect the Results from Incentive-Compatible Experiments? The Case of Non-GM and Country-of-Origin Premia in Canola Oil. *Canadian Journal of Agricultural Economics* 57: 205–221.~~

~~Yue, C., and C. Tong. 2009. Organic or Local? Investigating Consumer Preference for Fresh Produce using a Choice Experiment with Real Economic Incentives. *HortScience* 44 (2): 366–371.~~