The Effects of Honesty Oath and Consequentiality in Choice Experiments

Nathan Kemper
University of Arkansas
nkemper@uark.edu

Rodolfo M. Nayga, Jr.
University of Arkansas

Jennie Popp
University of Arkansas

Claudia Bazzani
University of Arkansas

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2016 AAEA Annual Meeting, Boston, Massachusetts, July 31-August 2

The Effects of Honesty Oath and Consequentiality in Choice Experiments

Abstract

Choice experiments are now one of the most popular stated preference methods used by economists. A highly documented limitation of stated preference methods is the formation of hypothetical bias in the estimation of consumers' willingness-to-pay (WTP) for a good or a service. Honesty oaths and consequentiality scripts are two *ex ante* approaches that show promise in their ability to reduce or eliminate hypothetical bias. We examine these approaches independently and together and measure their effectiveness by comparing the resulting WTP values. We also explore a potential connection between consequentiality, honesty oaths, and attribute non-attendance (ANA). We infer patterns of ANA resulting from our various treatments (i.e., consequentiality script only, honesty oath only, combined script and oath, inconsequential, and control) and examine the differences. Our results suggest that the combined *ex ante* approach of consequentiality script and honesty oath provided significantly lower WTP values than all other experimental treatments. Conditioning our data for both consequentiality and ANA resulted in significant improvements in model fit across all treatments. Results indicate that not accounting for ANA has important implications for welfare estimates. While we cannot fully explain the connection, the combination of the consequentiality script, honesty oath, and inferred ANA allowed us to better see the differences between respondents' attending attributes and those ignoring.

Acknowledgements: the authors express their gratitude for contributions made by Keith Chrzan on the experimental design, survey construction, data collection and advice on data analysis. We also thank Ellen Van Loo, Vincenzina Caputo, William Greene, Gregory Poe, and David Hensher and for helpful suggestions.

The Effects of Honesty Oath and Consequentiality in Choice Experiments

1. Introduction

Choice experiments are now one of the most popular stated preference methods used by researchers to elicit an individual's preferences for public and private goods. However, one limitation of choice experiments as a stated preference method is the formation of hypothetical bias in the estimation of consumers' willingness to pay (WTP) for a good or a service (Murphy et al., 2005). This bias is (simply) the difference between what a person *says* he/she would pay and what a person would *actually* pay (de-Magistris et al., 2013; Murphy et al., 2005). Because stated preference methods are so widely used, a great deal of academic research has been devoted to developing ways to mitigate the bias. Honesty oaths and consequentiality scripts are two *ex ante* approaches that show promise in their ability to reduce or eliminate hypothetical bias. Honesty oaths require participants to sign an oath promising that they will tell the truth. Consequentiality scripts are designed to increase participants' perception that their responses influence an outcome that they care about. Research on the honesty approach has demonstrated its usefulness in a number of disciplines (Loomis, 2014; de-Magistris et al. (2013); Jacquemet et al. (2009, 2010, and 2013). The literature on the use of consequentiality scripts in private markets in a discrete choice experiment setting is still developing. Consequentiality is also used as an *ex post* approach to calibrate models to adjust for (or eliminate) responses from participants who perceive the survey to have low or no consequentiality. Tests comparing consequential to inconsequential treatments have found that different responses are obtained from respondents in the inconsequential treatment (Carson, Groves, and List, 2014). This suggests that results obtained for the inconsequential case should not be used to make inferences about how the consequential group behaves.

A growing area of research has focused on attribute non-attendance (ANA) as a source of bias in choice experiments. Respondents may ignore some of the attributes while evaluating alternatives in a choice task (Hess and Hensher, 2010; Scarpa et al., 2009) and therefore respondents may not make the trade-offs between all the attributes, as theoretically assumed, due to ANA. Overlooking ANA in choice models can affect coefficient estimates, model fit and performance measures (Campbell et al., 2008; Carlsson et al., 2010; Hensher, 2014; Hensher and Rose, 2009; Scarpa et al., 2009). If an individual ignores an attribute, this implies that he/she is making no trade-off between the ignored attribute and another attribute. Thus, individual-level estimates resulting from models not accounting for ANA could contain errant calculations for marginal rate of substitutions and potentially misleading WTP estimates, particularly if price is one of the ignored attributes. Hence, conditioning models to account for ANA patterns is essential and these patterns can be inferred based on stated choices by respondents.

Our research examines the effectiveness of three approaches to mitigate hypothetical bias in a choice experiment: 1) honesty oath, 2) consequentiality script, and 3) combined approach. Effectiveness of these three approaches is measured against two control groups: 1) baseline control - a group given no honesty oath or consequentiality script, and 2) inconsequential control - a group given no honesty oath or consequentiality script but also informed that their responses would not be "consequential" in any way. Hence, our study employs a between-sample design where respondents are randomly assigned to one of the five groups (i.e., the two control groups, honesty oath group, consequentiality script group, and the combined approach group).

In this study we seek to address four general questions. First, do two common *ex ante* approaches result in lower WTP values over the controls (thus mitigating hypothetical bias)? Second, can we combine the two *ex ante* approaches in order to return even lower WTP values? Third, does conditioning our data for ANA improve our models and reduce bias in our data?

2. Hypothetical Bias in Stated Preference Surveys

There are several different types of stated preference approaches (Kling et al., 2012). Once of the most widely used is the discrete choice experiment (DCE). In a DCE, participants are asked to consider a product that is defined by several attributes and a no-choice alternative (Hensher et al., 2015). DCEs allow for the identification of the tradeoffs that each individual makes between attributes; when price is included it is possible to estimate marginal value of changes in the attributes. The DCE is a widely used method in valuing products and attributes (de-Magistris et al., 2013).

In the assessment of both private and public goods, researchers have observed a discrepancy between what a person indicates they would pay in the survey (hypothetical) and what a person would actually pay (non-hypothetical) (Champ et al., 1997; Harrison, 2006; Loomis et al., 2000). This well-known shortcoming of stated preference methods is known as hypothetical bias and its existence is demonstrated by a large body of empirical work (Harrison, 2006). While no theoretical approach in stated preference research can fully explain the existence of hypothetical bias (Mitani and Flores, 2010) it is clear that hypothetical values differ from real values (List and Gallet, 2001; Murphy et al., 2005; de-Magistris et al., 2013). Several approaches to reduce hypothetical bias have emerged in the literature. *Ex ante* approaches seek to reduce the bias by survey design while *ex post* approaches involve the calibration of WTP responses to remove the hypothetical bias from the stated WTP data.

2.1 Honesty Oath and Consequentiality

There is no consensus as to which approach is the best to correct for hypothetical bias. Some approaches are more useful for public goods while others for private goods; some are more practical than others (Loomis, 2014). Two straight forward *ex ante* approaches are honesty oaths and consequential scripts. Both approaches have shown promise in reducing and eliminating bias. Honesty oaths have been included in the study of private market goods. Consequentiality scripts have almost exclusively been used in the assessment of public goods. Little is known about the effectiveness of a consequentiality statement to "prime" participants to give more accurate WTP estimates in a DCE evaluating private goods; therefore, more space is devoted to consequentiality scripts than to honesty oaths in the following subsections.

2.1.1 Honesty Oaths

One hypothesis about the source of hypothetical bias is that due to the lack of economic incentives, respondents do not take the hypothetical task seriously or do not exert enough cognitive effort to provide accurate answers (de-Magistris et al., 2013). To address this issue, Jacquemet et al. (2009, 2010, 2013) used an oath which participants signed and promised to tell the truth and provide honest answers. The researchers hypothesize that when the participant makes a promise in a hypothetical setting they will be more inclined to provide an unbiased and more accurate answer. Kulik and Carlino (1987) concluded that when parents promised to give their children antibiotics they were more likely to actually administer the antibiotics. Joule et al. (2007) found that people who promised to use energy efficient light bulbs did so more than people who did not make a promise. Jacquemet et al. (2013) compared the oath to a cheap talk script and their results suggest the solemn oath improves the revelation of true preferences in both real and hypothetical contexts and outperformed cheap talk in reducing or eliminating hypothetical bias. The researchers argued that the oath can be used as a truth telling commitment device (Jacquemet et al., 2013); however, further research is needed on whether their results can be replicated in other contexts.

The honesty oath approach shows much potential in its ability to reduce hypothetical bias. Compared to cheap talk, which has been used with mixed results, research has shown honesty oaths to be more effective in reducing hypothetical bias. The oath is also simple and easy to implement. Participants are familiar with the language used as it is similar to oaths used in the US court system and it takes little time to read, understand and sign. It is easier to implement than another promising honesty approach developed by de-Magistris et al. (2013) which required participants to unscramble and rewrite 24 scrambled sentences containing honesty statements. By comparison, the honesty oath used by Jacquemet (2009, 2010, 2013) contained one sentence: “I undersigned swear upon my honor that, during the whole experiment, I will: Tell the truth and always provide honest answers.” Based on the practicality and effectiveness of the honesty oath, it will be included in our project.

2.1.2 Consequentiality (Ex Ante)

Carson and Groves (2007) argued that a hypothetical survey can yield more than hypothetical answers if the survey is perceived by participants to be consequential. The use of the *ex ante* consequentiality approach has been primarily applied in settings using voter referendum style dichotomous choice survey instruments evaluating public goods. The effectiveness of the approach in reducing hypothetical bias under these conditions is promising. Bulte et al. (2005) found that participants in a field experiment who were directly confronted about hypothetical bias behaved similarly to participants who were told that survey responses would be seen by policy makers. Vossler and Watson (2013) found that respondents who believed that survey results would not be considered by policy makers were less likely to vote in favor of a program.

Binary choice surveys have been the focus of research on consequentiality because of their perceived ability to be incentive compatible. Carson and Groves (2007) argue that as long as survey participants perceive some positive probability that a positive vote increases the likelihood that the actual project will be implemented, that the incentive properties of the survey hold. The ability of respondents’ to map the one-to-one relationship between their responses and actual policy outcomes is what makes a survey theoretically incentive-compatible. In binary choice experiments, the role of respondents’ beliefs about this mapping is seen as key to obtaining reliable estimates. This concept was termed the “knife-edge” by Herriges et al. (2010) who also found evidence to support the idea that participants who perceive that the survey outcome will affect implementation behave differently than those who do not. However, results differ on the how high the perception of consequentiality needs to be to trigger incentive properties in a binary choice experiment.

Carson and Groves (2007) and Herriges et al. (2010) found no differences in participants who reported perceiving the survey as “very likely” and “somewhat likely” to influence policy adoption. Vossler et al. (2012) concluded that a relatively higher level of perception of consequentiality is required to trigger incentive properties. Importantly, Vossler et al. (2012) also found evidence that economic incentives appear to be activated in a stated preference survey where financial consequences are remote and the payment mechanism is vaguely articulated. This conclusion offers hope for researchers using DCEs because such mapping of the relationship between responses given and actual policy (or market) outcomes are not always known by the researchers and therefore not described to participants.

The usefulness of a consequential approach in a multinomial choice setting remains largely unexplored. Multinomial choice experiments are popular among researchers because they allow for a more direct means of estimating attribute effects. Interis and Petrolia (2014) examined the effect of respondent perceptions of consequentiality on a multinomial choice stated preference survey and observed the “knife-edge” results in their data on preferences for wetland and barrier island restoration. The researchers concluded that neglecting to account for perceived consequentiality can lead to false conclusions. They state that consequentiality may be unlikely to aid incentive compatibility but that consequentiality is still a necessary condition for theory to be able to make predictions in a multinomial setting. Although Interis and Petrolia (2014) did inform respondents that the survey results would be shared with policy makers to help them make policy decisions, the study was not designed to address the ability of the *ex ante* consequentiality approach to “prime” respondents to provide accurate responses.

While the consequentiality approach appears to be effective at reducing hypothetical bias in the evaluation of public goods, its use in the evaluation of private goods is limited. Carson et al. (1997) demonstrated that the even under the dichotomous choice format, in the case of provision of a new private or quasi-public good, the survey is not incentive compatible. Incentive compatibility can only be restored for such goods if the binary choice is between two different forms of the good (Carson and Groves, 2007). Such a choice design results in a valuation question that represents a change in the good. Drichoutis et al. (2015) evaluated the effectiveness of a consequentiality script in a contingent valuation study of a private good with and without a fair labor label. They found their consequentiality script (modeled after the script used by Vossler and Watson (2013) and Vossler and Evans (2009)) to have no effect in mitigating hypothetical bias. They also found a cheap talk script to have no effect on bias. The researchers observed problems in using scripts verbatim from previous studies and that more research is needed to test the effectiveness of such scripts in reducing hypothetical bias in different contexts.

Herringes et al. (2010) characterized two necessary conditions under which researchers can expect respondents to answer truthfully. Respondents must believe that the results could influence an outcome they care about (*policy* consequentiality) and they must perceive that there is some probability that they will have to pay for the program via a tax (*payment* consequentiality). Together, the two conditions result in *strong consequentiality* and under these conditions respondents' dominant strategy is to answer truthfully. Attention should be given to both *policy* and *payment* consequentiality in the script design to trigger respondents' beliefs that their answers will be consequential in both senses. The consequentiality scripts used in recent research (e.g. Drichoutis et al., 2015; Vossler and Watson, 2013; Vossler and Evans, 2009) tend to focus on the *policy* consequentiality with little or no attention paid to the impact on the person's budget via associated taxes or fees, if the proposal is passed. This may limit the script's usefulness in reducing hypothetical bias. Including language that focuses participants on the financial implications of their responses may improve its performance. Many cheap talk scripts include (in addition to a description of hypothetical bias) language encouraging participants to consider what they are truly willing to pay and that their decision will impact their budget. Including similar language in a consequentiality script could be used to signify *payment* consequentiality to participants.

2.1.3 Consequentiality (Ex Post)

Consequentiality can also be used to calibrate the results from DCEs. In this *ex post* consequentiality approach, respondents are asked to rate the extent to which they believe their answers will be taken into account by some decision maker (typically policy maker or government agent) (Vossler and Watson, 2013). The responses to these questions can then be used to calibrate results by comparing "consequential" responses to "inconsequential" responses and, if a significant difference exists, the inconsequential responses are removed from the data. Carson et al. (2014) conclude that for a consequential question, the probability that the responses are taken into account in making a decision does not matter *as long as it is bounded away from zero*. They found that a different response as obtained from respondents at $p = 0$ (zero probability of affecting an outcome), in both the mean and variance of responses. This implies that only responses from participants indicating a zero probability ($p = 0$) should be removed in order to condition DCE data for consequentiality. This leads to a data set containing responses from those who believe their responses to have some positive probability ($p > 0$) of affecting some outcome. However, simply believing one's responses have an impact on a real outcome does not necessarily lead to more truthful responses (or incentive compatibles ones). Carson et al. (2014) also conclude that respondents will take advantage of transparent incentives that encourage misrepresentation. In our study, somewhat controversial labels are included in the DCE (relating to the genetically engineered content) and some respondents could misrepresent their *true* responses in order to attempt to influence some outcome of importance to them. For instance, if a respondent has a negative view of genetically modified foods and wanted to respondent accordingly, this respondent could ignore price and all other attributes in the experiment and respond solely to “cast their vote” for the non-genetically modified label. This would certainly be reflected in the individual’s strong preferences for the non-genetically modified label; however, this would also be considered non-attending of the other attributes and any conclusions about the other attributes based on this individual’s results would be erroneous. Numerous examples of how respondents may try and exploit what they see as transparent incentives can be imagined, including one that takes the exact opposite position in strong support of genetically modified foods. This highlights the importance of identifying and conditioning data for ANA. We can expect respondents to report truthfully when it is in their interest to do so; however, we can also expect that when it is not in their interest to do so, that respondents may give inaccurate and/or incomplete responses. It is also possible that some respondents may not fully understand attributes and therefore cannot give completely accurate responses.

3. Attribute Non-Attendance (ANA)

Strategies used by respondents in choice experiments play an important role in understanding the how individuals assess attributes associated with choice alternatives (Hess and Hensher, 2010; Scarpa et al., 2009; Erdem et al., 2015). There is a growing body of research focused on these issues that suggests that respondents may follow a number of different decision rules to simplify decisions (Hensher 2006; Scarpa et al., 2013). These heuristics ultimately result in non-attendance to certain attributes. Recently, ANA has become the focus of much research in order to better identify ANA and calibrate models to account for ANA. Self-reported statements of ANA have been included in surveys in order to condition models based on stated ANA (Hensher, 2006; Hensher and Rose, 2009; Islam et al., 2007). However, while asking respondents direct questions seems to indicate that some respondents do consistently ignore certain attributes, it is not clear whether researchers should rely on this information during model estimation (Hess and Hensher, 2010). There are problems with endogeneity by conditioning the modelled choice process on stated processing strategies (Hensher, 2008) and the same concerns about the quality of responses in the choice data extends to direct questions about decision-making heuristics. If stated measures of non-attendance are affected by respondent inaccuracies due to accidental or intentional misrepresentation, such measures would be uninformative and invalid. If this is indeed the case, the best course of action for researchers may be to use statistical methods of ANA inference.

ANA can be inferred through the estimation of analytical models (Van Loo et al., 2015 and Bello and Abdulai, 2016) and is often based on latent class or mixed logit models (Hess and Rose, 2008; Hess and Hensher, 2010; Caputo et al., 2013 and Scarpa et al., 2013; Collins and Hensher, 2015). The equality constrained latent class method imposes specific restrictions on the utility functions for each class of respondents by constraining some coefficients to zero for selected attributes respective classes (Caputo et al., 2013; Hensher and Greene, 2010; Scarpa et al., 2009; 2013). Hess and Hensher (2010) suggest inferring ANA through the use of mixed logit models which are used to first derive individual-level estimates of coefficients and variance which are then used to examine respondent specific coefficient of variation in order to identify large “signal-to-noise” ratios and thereby infer ANA. Scarpa et al. (2013) compared the stated methods to both latent class and mixed logit methods of inferring ANA and concluded that it is not possible to identify which of the approaches *best* accounts for ANA, but that ignoring ANA behavior in choice experiments has numerous consequences for welfare estimates including WTP measures.

4. Experimental Design and Methods

4.1 Experimental Design

The data were collected through a national, web-based choice experiment survey built using the software package Sawtooth Software (www.SawtoothSoftware.com) and collected by Survey Sampling International (SSI) (www.SurveySampling.com) using their nationally representative consumer panel. The panel consisted of 2,535 participants who were the primary grocery shoppers for their households randomly placed into one of five treatments with approximately 500 participants per treatment. The sample frame from SSI is balanced by socio-demographic characteristics and we also took in to account the four main US Census regions to regionally balance the sample across the US. The experiment consisted of two tasks. First, respondents participated in a DCE where they made choices between poultry products differentiated by various labels regarding the presence of genetically modified ingredients, production location, and carbon footprint. Second, respondents were asked a series of survey questions regarding their perception of the consequentiality of the DCE as well as other questions relating to food preferences and demographic data. The study uses a between-subject design where respondents participate in only one of the treatments and because our target population is consumers and not students we have a non-standard subject pool (Harrison and List, 2004).

4.2 Choice Set Design

Boneless skinless chicken breast was chosen for use in the DCE for four reasons. First, the overarching project goal was to evaluate market opportunities for soybean farmers. Second, only recently have meat and poultry products used non-GM label statements. Third, boneless skinless chicken breast is a widely consumed product in the US. Fourth, the product is sold in packages that could carry a non-GM label. Two complementary labels were included in the study: local production of both birds and feed and carbon footprint.

Table 1 shows the choice experiment attributes and levels (with corresponding effects coding). Effects coding was used due to the benefits provided when there are potential interactions between two categorical variables (such as local and carbon footprint). Effects coded data provide reasonable estimates of both main effects and interaction effects whereas data dummy coded data provide only simple effects (the effects of one variable at one level of the other variable) (Bech and Gyrd-Hansen, 2005). In this study, the ability to clearly examine and distinguish the main and interaction effects of all attribute levels is of great importance. Price has four levels chosen to reflect 2015 nominal prices found across US supermarkets. Prices were sampled from retail outlets of both brick and mortar stores and online retailers. USDA price reports for chicken were also consulted (USDA ERS, 2015). A goal of this study was to determine consumers' preferences for chicken breast meat carrying a Non-GMO Project Verified label. Therefore, the second chosen attribute was genetically modified (GM) content which had three levels: 1) no information, 2) Non-GMO Project Verified, and 3) "this product contains genetically engineered ingredients". Permission was granted by the Non-GMO Project to use their logo, statement and label in our hypothetical experiment (www.nongmoproject.org). The GM labels were selected as they are currently all valid labeling options under the US system of voluntary labeling. However, the authors acknowledge that while companies can label their foods as "containing GM ingredients" under the current system, in reality possibly all of the labeling currently in the marketplace focuses on the non-GMO attribute. With the recent developments around GM labeling in the US we included the "contains GM" language, in part, to gauge how consumers may respond to such language if it were to appear on products in the future (possibly due to new state or federal regulations). We also sought to examine consumers' preferences for two additional sustainability labels: carbon footprint and local production. The third attribute was carbon footprint which had 4 levels: no information, low, medium and high carbon footprint (values of CO_2 in Table 1). The CO_2 levels followed those used by Van Loo, et al.

(2014). The fourth and final attribute was local production which was defined by the birds and feed being grown in the respondent's own state. The "local" attribute had two levels: no information and "birds and feed grown in your state."

Each respondent was presented with eight choice tasks, where each choice task included a no-buy option and two experimentally-designed options (see Figure 1 for an example choice task.). The allocation of attribute levels to alternatives was designed following Scarpa et al. (2007), using a sequential Bayesian design to minimize the Db error. Different design phases were conducted. In the first stage an orthogonal design (Bliemer and Rose, 2010) was used for a pilot survey on 250 respondents. The data from the pilot study were then used to estimate a MNL model whose coefficient estimates were then implemented as Bayesian priors for the data collected in the first wave. All designs involved 32 choice tasks arranged in four blocks of eight tasks each and were obtained and evaluated using Sawtooth Software and Ngene version 1.1.2 (Choice Metrics, 2012).

4.3 Treatment Descriptions and Hypotheses

Concerning our two questions regarding mitigation of hypothetical bias, we specify and tested a series of hypotheses based on the experimental treatments. To answer our first two questions regarding the ability of two *ex ante* approaches and their combined use to result in lower WTP values over the controls, we tested six hypotheses. We also tested four additional hypotheses comparing the inconsequential control to the baseline control and the *ex ante* approaches against each other.

Treatment 1 used a consequentiality script (CS). Herringes et al. (2010) used *policy* and *payment* consequentiality to describe two areas of consequentiality to emphasize in the evaluation of a public good. We used similar concepts and adapted terminology for use in a market setting with a private good. Specifically, we developed a script that suggested to respondents to believe that their answers could influence future product offerings (*product* consequentiality) and that their choices should reflect their true preference and that they should consider that their budget for other purchases will be reduced by the same amount as the dollar amount in their choice (*budget* consequentiality). By emphasizing both *product* and *budget* consequences to participants, we hypothesize that this would induce them to put more effort into the choice task. Our CS was adapted for context from Drichoutis et al. (2015), Vossler and Watson (2013) and Vossler and Evans (2009) to emphasize *product* consequentiality. Our additional language to emphasize *budget* consequentiality was adapted from two cheap talk scripts used by List (2001) and de-Magistris et al. (2013) and reads as follows:

"We would like to inform you that the survey results will become available to producers, manufacturers, and retailers of agricultural products as well as to policy makers and the wider general public of consumers. This means that this survey could affect the decision of producers, manufacturers, and retailers to introduce new products or make changes to current products based on your responses. Because of the importance of the survey, we ask that before selecting an option in each choice question, please try to think the same way you would if you really had to pay for the product and take it home. Sometimes in experiments like this one, people choose a product with a price that represents their best guess of what the product is really worth. But, when people actually have to spend their own money, they may not make the same choice because they take into account the limited amount of money they have. Just like in a real retail setting, please take into account how much you really want to spend your own money on the product and consider the impact on your budget."

We test the hypotheses that individuals who read the CS indicate a lower WTP than those in the control group who are not exposed to the script and those in the inconsequential control who are informed directly that their responses will not be used to make product, policy or pricing decisions.

$H0_1 : (WTP^1 - WTP^5)\ 0$, and
$H1_1 : (WTP^1 - WTP^5) < 0$

$H0_5 : (WTP^1 - WTP^4) = 0$, and
$H1_5 : (WTP^1 - WTP^4) < 0$

If $H0_1$ is rejected, we might confirm that introducing the CS in the hypothetical CE reduces hypothetical bias because the WTP values using CS would be lower than in the baseline control (BC). If $H0_5$ is rejected, we might confirm that introducing the CS in the hypothetical CE reduces hypothetical bias because the WTP values using CS would be lower than in the inconsequential control (IC).

Treatment 2 used an honesty oath (HO) based on the oath used by Jacquemet et al. (2009, 2010, and 2013) and reads as follows:

"I undersigned swear upon my honor that, during the whole experiment, I will: Tell the truth and always provide honest answers."

We test the hypotheses that individuals who sign the HO indicate a lower WTP than those in the control group not exposed to the oath and those in the inconsequential control.

$H0_2 : (WTP^2 - WTP^5) = 0$, and
$H1_2 : (WTP^2 - WTP^5) < 0$

$H0_6 : (WTP^2 - WTP^4) = 0$, and
$H1_6 : (WTP^2 - WTP^4) < 0$

If $H0_2$ is rejected we might confirm that introducing the HO in the hypothetical CE reduces hypothetical bias because the WTP values in the HO would be lower than in the BC. If $H0_6$ is rejected we might confirm that introducing the HO in the hypothetical CE reduces hypothetical bias because the WTP values using CS would be lower than in the inconsequential control (IC).

Treatment 3 combined the use of the CS and HO (CSHO). The CS script was shown first followed immediately by the oath. Because the HO and CS could induce honest behavior from different angles, the approaches may complement one another. The HO may influence participants who have just read the CS take the script more seriously, thereby increasing the consequentiality of the choice experiment.

We test the hypotheses that individuals in Treatment 3 who read the CS followed by the HO indicate a lower WTP than those individuals in the control group and in the inconsequential control.

$H0_3 : (WTP^3 - WTP^5) = 0$, and
$H1_3 : (WTP^3 - WTP^5) < 0$

$H0_7 : (WTP^3 - WTP^4) = 0$, and
$H1_7 : (WTP^3 - WTP^4) < 0$

If $H0_3$ is rejected we might confirm that combining the HO and the CS in the hypothetical CE reduces hypothetical bias because the WTP values in treatment 3 would be lower than in the hypothetical CE. If $H0_7$ is rejected we might confirm that CSHO in the hypothetical CE reduces hypothetical bias because the WTP values using CSHO would be lower than in the inconsequential control (IC).

We also test the hypotheses that individuals in Treatment 3 indicate a lower WTP than individuals in either Treatments 1 or 2.

$H0_9 : (WTP^1 - WTP^3) = 0$, and
$H1_9 : (WTP^1 - WTP^3) < 0$

$H0_{10} : (WTP^2 - WTP^3) = 0$, and
$H1_{10} : (WTP^2 - WTP^3) < 0$

If $H0_9$ is rejected we might confirm that combining the HO and the CS in the hypothetical CE does not significantly outperform the CS alone. If $H0_{10}$ is rejected we might confirm that combining the HO and the CS in the hypothetical CE does not significantly outperform the HO alone.

Finally, we test the hypothesis that individuals in treatment 1 who read the CS indicate a lower WTP than those individuals who sign the HO in treatment 2.

$H0_8 : (WTP^1 - WTP^2) = 0$, and
$H1_8 : (WTP^1 - WTP^2) < 0$

If $H0_8$ is rejected we might confirm that the CS returns lower WTP values than the HO.

4.4 Conditioning Data for Consequentiality and ANA

We included a consequentiality survey question after the CE. The question was adapted from Drichoutis et al. (2015) and Vossler and Watson (2013) and uses a five point scale and appears as:

“To what extent do you believe that your answers in this survey will be taken into account by producers, manufacturers, retailers, and policy makers?”

A response of “1” indicates “not taken into account” and a “5” indicates “definitely taken into account”. Using the results of these questions we first test for differences between our treatments and examine any effect the CS has on respondents’ perception of the consequentiality of the survey. If the CS is effective, we should observe a statistically significant difference between our treatments that include the CS (CS and CSHO treatments) as well as in our inconsequential treatment (IC). We expect treatments using the CS to have the lowest number of respondents' perceiving the survey to have no consequences.

We also use responses to the consequentiality question as an *ex post* approach to condition our data. Carson, Groves, and List (2014) provide strong evidence that those respondents who perceive their responses to have zero probability of affecting some outcome (*p*=0) should not be used to make inferences about how the rest of the respondents behave. Models are first built to include these "inconsequential" respondents. Next, (p=0) respondents are removed from the analysis and models are then estimated.

The connection between ANA and consequentiality has not been identified in the literature, but we explore the potential of a connection here. The consequentiality approach attempts to induce respondents to provide more thoughtful answers in the CE. If our CS succeeds, we should expect respondents to put more cognitive efforts towards making decisions in our CE. However, *thoughtful* answers do not equate to *honest* answers. Could inducing respondents to give more thought to affecting some potential outcome such as new product development, pricing strategies, their own budgets or GM labeling policies actually influence some respondents to give misleading responses? Carson et al. (2014) concluded that respondents *will take advantage of transparent incentives that encourage misrepresentation*. Because in our study we are encouraging respondents to consider outcomes relating to controversial topics, namely the genetically engineered content of foods, we would expect that some respondents may perceive some ulterior motive on our behalf (possibly that we are pro- or anti- GMO for instance). We may then be indirectly inducing these respondents to misrepresent their *true* responses in an attempt to influence some outcome of importance to them. Rather than inducing misrepresentation, we may induce adoption of some heuristic in order to influence an outcome, such as always selecting the non-GMO label in order to express one's "vote" for non-GMO (or against GM). If we also assume that the most controversial attributes are the ones gaining the most attention from respondents, then they may be ignoring other attributes such as carbon footprint and local. This highlights the importance of identifying and conditioning data for ANA. We can expect respondents to report truthfully when it is in their interest to do so; however, we can also expect that when it is not in their interest to do so, that respondents may give inaccurate and/or incomplete responses. There are also likely some respondents who do not fully understand all of the labels and therefore cannot fully attend all attributes.

Identifying ANA regardless of the source of the ANA is well-documented as a necessary step in analyzing choice data in an experiment such as ours. The choice literature emphasizes the importance of taste heterogeneity. Therefore, we use the mixed (random parameters) logit (MXL) approach with error components to be able to evaluate ANA in the context of models able to address random taste variation (Train, 2003). To identify patterns of ANA we follow the procedures proposed by Hess and Hensher (2010) using MXL models. Their method is based on the coefficient of variation of individual specific posterior means and variances. Huber and Train (2001) demonstrate the derivation of conditional distributions of taste parameters from the estimated population parameters. Assume that respondent n has a normally distributed coefficient for attribute k, then $\beta_{kn} \sim N(\mu_{kn}, \sigma^2_{kn})$ where μ_{kn} is the estimated mean and σ^2_{kn} the variance. The coefficient of variation (CV) $\kappa_{kn} = \sigma_{kn}/\mu_{kn}$ is then interpreted as the "noise-to-signal" ratio on the variation relating to taste intensity for attribute k as evidenced by the individual's responses in the choice tasks (Scarpa et al., 2013). If the noise-to-signal ratio is high (above 2 in our case, CV>2), then the respondent-specific normal distribution is over-dispersed and the pattern of choice is consistent with the respondent not attending to attribute k in their choices. Hess and Hensher (2010) use the CV value of 2, so that a respondent n is considered to not be attending attribute k if their estimated value of $\kappa_{kn} > 2$. The choice of using the CV value 2 is based on the observation that normal distributions with ratios higher than 2 are over-dispersed (Scarpa et al., 2013). The sample proportion of ANA is then obtained by aggregating these values.

4.5 Econometric Methodology

Respondents' preferences and WTPs were analyzed using a discrete choice framework consistent with random utility theory (McFadden, 1974) and Lancaster consumer Theory (Lancaster, 1966). A Mixed (Random Parameters) Logit (MXL) model with correlated errors and with error components was used to estimate preferences and WTP. The utility function is specified as follows:

$$(1)\ U_{ijt} = \text{NONE} + \beta_1 PRICE_{ijt} + \beta_2 \text{NGE}_{ijt} + \beta_3 \text{GME}_{ijt} + \beta_4 \text{LOE}_{ijt} + \beta_5 \text{MDE}_{ijt} + \beta_6 \text{HIE}_{ijt} + \beta_7 \text{LCE}_{ijt} + \eta_{\text{ijt}} + \varepsilon_{ijt}$$

where *i* is the individual respondent, *j* refers to three options available in the choice set (Product A, Product B, and None) and *t* referring to the number of choice situations. The alternative-specific constant (NONE) is dummy coded taking the value 1 for the no-buy option and 0 otherwise. *PRICE* is a continuous variable represented by the four experimentally designed price levels ($2.99, $6.99, $10.99, $14.99). The non-price attributes Non-GMO (NGE), Contains Genetically Engineered Ingredients (GME), Low Carbon Footprint (LOE), Medium Carbon Footprint (MDE), High Carbon Footprint (HIE), and Local Production (LCE) are effects coded variables taking the value 1 if the product carries the corresponding labels, the value of -1 if the absence of the label, and 0 for the no-buy option. η is a zero-mean normally distributed respondent-specific error component shared by the two hypothetical alternatives reflecting a purchase decision and is absent in the utility of the no-buy (NONE) alternative. ε_{njt} is an unobserved random term that is distributed following an extreme value type-I (Gumbel) distribution i.i.d. over alternatives.

The common approach to estimating equation (1) is to assume price has a fixed coefficient. This is a widely accepted and practiced specification (Layton and Brown, 2000; Lusk and Schroeder, 2004; Revelt and Train, 1998). Fixing the price coefficient ensures that the estimated WTP will be normally distributed and all respondents will have a negative price coefficient. However, as pointed out by Scarpa, Thiene and Train (2008), this restriction is counter-intuitive as there are very good theoretical reasons as to why response to price should vary across respondents. Particularly for analyses such as ours where we specifically must look at individual-level coefficients and standard deviations in order to identify patterns of ANA, this restriction of a fixed price coefficient is incompatible. By assuming a fixed price coefficient, we would thereby assume that all respondents fully attend the price attribute. In a hypothetical study such as ours where the overarching purpose is to study methods to mitigate hypothetical bias, such a simplifying assumption is even more concerning. Therefore, in our study we allow all parameters to be random, thus allowing us to identify patterns of ANA in our data and condition new models based on these patterns[1]. The MXL models specified by equation (1) were used to infer ANA and compare model fits between specifications and treatments.

Scarpa, Thiene and Train (2008) found that estimating WTP directly using WTP space reduced the incidence of large WTP values and allowed for greater control in specifying the distribution of WTP. Following their method, to estimate WTP to test our hypotheses, we also specify our utility in WTP space rather than preference space as in equation (1). Our utility function is therefore re-written as:

$$(2)\ U_{ijt} = \alpha[\theta_1 NONE + N\_PRICE_{ijt} + \theta_2 NGE_{ijt} + \theta_3 GME_{ijt} + \theta_4 LOE_{ijt} + \theta_5 MDE_{ijt} + \theta_6 HIE_{ijt} + \theta_7 LCE_{ijt} + \eta_{ijt}] + \varepsilon_{ijt}$$

where $\theta_i = \beta_i / \alpha$ are already the WTP estimates. N_*PRICE* is a continuous variable representing our four price levels multiplied by -1 in order to facilitate WTP calculations with the correct sign. Following de-Magistris, et al. (2013), to test our hypotheses given this new utility specification, we pool data for the two treatments involved in the respective hypothesis and then specify an extended utility with the

[1] With price a random parameter rather than fixed, generating mean WTP values for a test such as the combinatorial test suggested by Poe, Giraud and Loomis (2005) was not appropriate. Rather, we estimated models in WTP space in order to estimate mean WTP values and tested our hypotheses accordingly in WTP space. However, as a robustness test of our data we also specified MXL models with a fixed price coefficient and used the Poe combinatorial method to test our hypotheses. These results are discussed briefly in the results section below.

appropriate set of treatment dummy variables depending on the hypothesis. Our extended utility function appears as follows:

$$(3)\ U_{ijt} = \alpha[(\theta_1 NONE + N\_PRICE_{ijt} + \theta_2 NGE_{ijt} + \theta_3 GME_{ijt} + \theta_4 LOE_{ijt} + \theta_5 MDE_{ijt} + \theta_6 HIE_{ijt} + \theta_7 LCE_{ijt}) + \delta_1 (NGE_{ijt} \text{ x } tr) + \delta_2 (GME_{ijt} \text{ x } tr) + \delta_3 (LOE_{ijt} \text{ x } tr) + \delta_4 (MDE_{ijt} \text{ x } tr) + \delta_5 (HIE_{ijt} \text{ x } tr) + \delta_6 (LCE_{ijt} \text{ x } tr) + \eta_{ijt}] + \varepsilon_{jt}$$

where *tr* is coded 1 for the first treatment in the analyzed hypothesis and 0 otherwise. For each of our 10 hypotheses, we specify one extended utility function and thus we used 10 *tr* dummy variables. The signs and significance of the estimated δ enable us to test differences in marginal WTP between the two treatments in the hypothesis to be analyzed.

5. Results

Tables 2 and 3 report the sample characteristics of the 2,535 respondents in the five treatments. Each respondent completed eight choice tasks with three choices per set for a total number of 60,840 observations (around 12,100 observations per treatment). Importantly, we also tested if there were differences in socio-demographic profiles across treatments using a chi-square test. The results of this test suggest that we cannot reject the null hypothesis of equality between characteristics across treatments and therefore our randomization was successful in providing a balanced sample across the five treatments (see table 2).

For the preference space models, we estimated equation (1) using a MXL with correlated errors and variance enhancing error components where price and all effects-coded attribute level variables are considered random following a normal distribution[2]. Estimations were conducted using NLOGIT 5 using 1,000 Halton draws to provide more accurate simulation for the random parameters (Train 1999).[3]

5.1 Consequentiality Results

Table 4 lists the number of respondents by treatment that fall into the "consequential" (*p*>0) and "inconsequential" (*p*=0) categories. As expected, treatments 1 (CS) and 3 (CSHO) have the lowest number of inconsequential respondents with 5.3% and 5.5%, respectively. The highest level of such respondents is also as expected in treatment 4 (IC). The null hypothesis of equality in rates of stated consequentiality across treatments is rejected for respondents stating that their responses were inconsequential. As shown in table 4, the overall numbers of respondents reporting that the results are inconsequential is quite small (26 for CS and 40 for IC) and models conditioned solely for these respondents, by removing their choices from the data, while important theoretically (see Carson et al., 2014) are practically not likely to have a substantial impact on model fit statistics. Because of the potential connection between the *perception* of consequentiality and the non-attendance of attributes, conditioning for consequentiality is a key step along the path to overall model improvement once data are conditioned for ANA, as our results will demonstrate. Table 5 reports the results from the baseline models across the five treatments and table 6 reports the models after being calibrated using the consequentiality data to remove (*p*=0) respondents' data. There are few changes in model fit statistics

[2] Numerous versions of the MXL models were estimated, using of normal, lognormal, and constrained triangular and combinations of these distributions. Models were also estimated with independently distributed as well as correlated coefficients and both dummy coded and effects coded models were used. The identification of ANA and how respondents were allocated to the attending and ignoring groups were comparable for the most part, with some combinations returning strikingly different allocations. Our search criteria sought to identify the models which 1) had the best fit to our data and 2) allocated respondents to ANA groups in ways that were sound both theoretically and rooted in common sense. For illustration purposes, we limit the results to the model using independent normal distributions for the random coefficients. Results from other models are available on request.

[3] Following Hensher and Greene (2003) all MXL models were estimated using 25, 50, 150, 250, 500, 1000 and 2000 draws to identify the number of draws required to produce stable results. Shuffled Markov-Chain draws and Halton draws were compared for use in simulations and returned similar results. Stable results were obtained at 1000 Halton draws and thus we adopted this for all of the models presented here.

and coefficient estimates across the models. However, we found that the models conditioned first for consequentiality performed better in the identification of ANA.

5.2 ANA Inference Models

With the data calibrated for consequentiality, another set of MXL models were estimated in order to identify patterns of ANA in our data. These models are presented in table 6. Several of our attribute coefficients in our original models (table 5) are not significant and have unexpected signs. Of particular concern is the carbon footprint attribute levels. Low carbon footprint (LOE) is positive and significant in only one treatment (CSHO) while the medium and high carbon levels have issues with preference reversal[4]. For example, the medium carbon footprint coefficient is negative in treatments 2 and 3 but positive in treatments 1, 4 and 5. The high carbon footprint coefficient has similar preference reversals but for different treatments. However, because these attribute levels are not significant (except medium in treatment 3) it is not clear what is truly causing the reversal or if there actually is a reversal. This could also signal that additional heterogeneity exists in our data that is not accounted for by the correlated coefficients and error components in our MXL models. ANA may also be responsible for the remaining individual-level differences.

Using the estimated individual-level coefficients and standard deviations, we used the models in table 6 to identify ANA in our data[5]. Table 7 shows the results of the inferred ANA using the noise-to-signal ratio of 2. We used the chi-square statistic to test the null hypothesis of equality in rates of ANA inferred across treatments. The null hypothesis is rejected for all attributes except price, indicating that rates of non-attendance were statistically different across the treatments. Treatments 1, 2 and 4 had the

---

4 One would not expect low carbon to have a positive coefficient (+), then medium to have a negative coefficient (-) and high carbon to then have a positive one (+). Such preference reversal may indicate ANA due to a lack of understanding of the carbon footprint labels or possible ignoring due to indifference.

5 We identified ANA using models conditioned for consequentiality and not conditioned for consequentiality, using a variety of distributional assumptions. We found that the best *final* models, as presented in table 8, were found when the sequence was first conditioning data for consequentiality and then identifying ANA.

lowest rates of non-attendance to the price attribute. These were the treatments CS, CSHO and IC. Although the differences between these treatments is not significant, it is interesting that the treatments mentioning consequences appear to increase the attendance to price. On the surface, it is puzzling that the IC treatment would have a lower rate of non-attendance to price than the baseline treatment. In the IC treatment, respondents are specifically instructed that their choices will not have an impact on their budget. Perhaps the mere mention of one's own personal budget induced greater attending to price.

Possibly of greater importance than the number of respondents not attending price, are what the parameters of conditional distributions show about respondent ignoring strategies from our data. We used the CV to incorporate uncertainty in conditional distributions as Hess and Hensher (2010) and allocated respondents as ignoring attributes with a CV greater than 2. Figure 2 shows a plot of the CV values for price coefficients in the five treatments. As shown in table 7, treatments incorporating consequentiality have the lowest rates of ANA to price. What is interesting about figure 2 is the relatively shorter scale for the CV values for treatment 3 (CSHO). Note the differences in the CV axis values; the CSHO treatment returned markedly lower noise-to-signal ratios than all other treatments. So while the CSHO treatment did not result in significantly lower numbers of ignoring respondents, it appears to have more effectively eliminated additional "noise" from respondents, likely due to ANA.

Aside from price, the rates of ANA were significantly different across the treatments, although no treatment is the clear "winner" in terms of the lowest rates of ANA. In terms of evaluating the effect of consequentiality on ANA it is not the rate of ANA that is important. Rather, the goal is to *reveal truthful preferences* and if consequentiality is successful at inducing such revelations, the truth may be revealed as higher *or* lower rates of ANA compared to the baseline. What is important is that we do test a difference in the rates of inferred ANA across treatments. Following Scarpa et al. (2013), we conditioned our data for ANA rather than attribute level non-attendance. Theoretically, ignoring any level of an attribute indicates the ignoring of the full attribute. If we do not make this assumption, we must allow for the possibility that a respondent could attend, for example, the Non-GMO attribute level but not the no information or GM attribute levels. A respondent may strongly *prefer* the Non-GMO attribute level but this is not the same as ignoring the levels she does not prefer. If a respondent with such a strong preference for Non-GMO uses the choice heuristic to focus exclusively on this attribute level alone, she would be considered ignoring the GM Content attribute because she is ignoring 2 of the 3 levels[6]. Figure 3 shows the patterns of ANA by treatment used to condition our data. The next test is whether conditioning our data based on the inferred ANA leads to better models.

5.3 Models Conditioned for ANA

Once data were conditioned for ANA a new set of models were estimated for our five treatments. Table 8 presents the results of the five MXL models. Model fit statistics indicate significant improvements over the models not conditioned for ANA. The best overall model in terms of fit statistics, using both the Bayesian information criterion (BIC) and Akaike information criterion (AIC) is the treatment 3 model which used the combined CSHO approach. Additionally, all coefficients and standard deviations in this model (and only this model) are significant and have the expected signs. Although the results for the models for treatments 1 and 2 (CS and HO) indicate that conditioning for ANA led to significant improvements in model fit, there is still potentially some heterogeneity unaccounted for in these models possibly due to ANA. The results in table 8 provide some, albeit limited, evidence that combining the CS and HO *ex ante* approaches can lead to better models based on the model fit statistics shown when combined with data conditioned for ANA. When examining the results from tables 6 and 7, we can see that the CSHO treatment is consistently the best model in terms of data fit; however, these models still contain coefficients which are not significant.

[6] We compared models conditioned for ANA and ALNA and found the ANA outperformed the ALNA models.

To further test the effectiveness of our MXL models in identifying patterns of ANA, we specify five models using only the "ignoring group" data. These are the individual-level data identified by our MXL as having large noise-to-signal ratios (CV>2). If our MXL models have done an effective job at identifying patterns of ANA, these ignoring group models should have coefficient estimates that are in opposition to theoretical and common sense expectations. Table 9 presents the results of these models. The first interesting result of these models is that price is positive and significant with an insignificant standard deviation. This seems to indicate that the respondents in the ignoring group were not adequately attending price. Other interesting findings include the negative preferences for low carbon and positive preferences for high carbon. This suggests that either some of our respondents did not understand the carbon footprint label or that for some reason respondents were misrepresenting their real preferences for the carbon labels. The no-buy (NONE) alternative specific constant is negative across the five treatments, which is theoretically sound; however, these size of the no-buy coefficients are substantially smaller than the models in tables 6, 7 and 8.

5.4 Willingness-to-Pay Space Models

Next, we compare the WTP results to further evaluate the effects of consequentiality, honesty oaths and ANA on marginal WTP (mWTP) values. Train and Weeks (2005) suggest that it is important to recognize that the scale parameter in many situations clearly does vary randomly over observations. Holding price fixed in order to estimate WTP errantly ignores variance in price across individuals which can lead to erroneous interpretation and policy conclusions. In the context of evaluating methods to mitigate hypothetical bias, constraining the price coefficient (when it indeed varies) will falsely attribute the variation in price to variation in WTP. Therefore, we re-parameterize our models such that the parameters are the marginal WTP for the attributes. The results of our models in WTP space are shown in table 10. The results indicate that coefficient (WTP) estimates are significant in all five treatments except the medium carbon attribute level in treatments 1,2,3 and 5 and high carbon in 3 and 4.

The lowest WTP estimate for the Non-GMO attribute are found in treatment 3, the CSHO treatment ($3.62/lb premium), and the highest WTP value in the baseline control ($7.01/lb). The combined treatment returns the lowest WTP values compared to the baseline control for the GM label ($2.89/lb to avoid GM), the low carbon footprint label ($0.93/lb) and local production label ($1.27/lb). Because the medium and high carbon footprint labels are not statistically significant across all models, little is gained by comparing their values here. The results provide further evidence that the combined use of CSHO has the potential to produce substantially lower mWTP estimates.

Tables 11 and 12 show the results of our hypotheses tests to examine the statistical differences between our treatments' mWTP values. Hypothesis 1 tested the difference between treatments 1 and 5 (CS vs BC) and the results indicate that the CS produced significantly lower WTP values for the Non-GMO attribute as well as the GM attribute. Since the sign of the WTP values are negative, a positive coefficient on the hypothesis test indicates a relatively lower WTP value. Hypothesis 2 tested the HO (treatment 2) against the BC (treatment 5) and the results indicate that the HO also results in significantly lower mWTP values compared to the BC; however, the reductions are not as substantial as with the CS. Hypothesis 3 compared the combined approach treatment 3 (CSHO) to the BC and the results indicate the highest level of reduction in WTP compared to the base. In addition, the CSHO approach also significantly lowers the WTP for the local production attribute. The hypotheses testing CS, HO and CSHO against the IC (hypotheses 5,6 and 7) offer similar results and support the general finding that the CSHO is the most effective of the three approaches at reducing mWTP values. However, due to the lack of significance of the medium and high carbon footprint attribute levels across many of the models, comparison of the mWTP values of these attributes is not appropriate.

5.5 Robustness Tests

Above, we note the limitations associated with holding the price coefficient fixed in order to ease the estimation of WTP and despite these limitations, treating price as such is common practice. In order to check the robustness of our data, we also specified MXL models in preference space with a fixed price coefficient. Based on the estimated coefficients from these models, we calculated mWTP for each attribute. Then, we again tested our hypotheses, this time using the combinatorial approach by Poe et al. (2005) to compare differences between mWTP estimates in the different treatments. The test requires the generation of a distribution of 1,000 WTP estimates which was carried out using the statistical software package R (R Core Team, 2013) using the Kinsky and Robb (1986) bootstrapping method. Coefficients and covariance matrices were estimated in NLOGIT 5 then analyzed in R. For the random draws we used a Bayesian estimator (James-Stein-type shrinkage estimator in the R package ‘corpcor’) in order to return a positive definite and well-conditioned covariance matrix across all treatments (Schäfer and Strimmer, 2005; Schäfer et al., 2015). The James-Stein estimator always improves upon the total mean square error (sum of expected errors of each component) and allows any particular component to improve for some parameter values and deteriorate for others. For this reason, such as estimator is preferred when three or more parameters are estimated.

Tables 13 and 14 summarize the mWTP estimates and hypotheses tests using the Poe et al. (2005) combinatorial approach. The results are similar to our WTP space models results. The significant *p*-values in tables 13 and 14 are in bold and italics and indicate that treatment 3 (CSHO) results in significantly different mWTP estimates for the Non-GMO, GM, and Local attribute levels when compared to the BC (hypothesis 3) and the Non-GMO, GM, and High Carbon attributes levels when compared to the IC (hypothesis 7). Our WTP space models did not detect a significant difference in the High Carbon attribute levels comparing CSHO to the IC.

6. Conclusion

The original intent of this paper was to examine the effect of honesty oath and consequentiality script on mitigating hypothetical bias in stated preference methods using choice experiments. A highly documented limitation of stated preference methods is the formation of hypothetical bias in the estimation of consumers' WTP for a good or a service. This bias is usually assumed to put upward pressure on WTP estimates from hypothetical choice experiments. However, our research quickly developed into a study of the potential relationship between our *ex ante* and *ex poste* bias mitigation methods and ANA. For the purpose of discussion, table 15 compares the marginal WTP estimates from the aggregate multinomial logit (MNL) models from the baseline data (before conditioned for consequentiality and ANA), the MNLs from the ANA conditioned data, attending group (AA MNL), and the MXLs from the ANA conditioned data, attending group (AA MXL). These results are presented here to demonstrate a possible flaw in the conventional perception of how ANA affects WTP values. As seen in table 15, the WTP values from the conditioned data (AA MNL models) are substantially *larger* than those from the unconditioned data models. As noted by Scarpa et al. (2013), one would expect that addressing ANA would provide lower mWTP values. Because the models are conditioned to remove responses from those participants who are not adequately attending price, logically, the values would be expected to be lower. However, in a purely hypothetical setting, our results indicate that accounting for ANA may actually lead to significantly higher mWTP values. Also of note in table 15 is that treatment 3 (CSHO) returns the lowest mWTP values for the significant attributes in our models (Non-GMO, GM, and Local) in the baseline MNL, AA MNL and AA MXL. This provides further evidence that the combined approach of CS and HO could be acting on different sources of hypothetical bias and, once combined in practice, could lead to more thoughtful and more truthful answers from respondents. Figure 3 which compared the noise-to-signal ratios across treatments also supports this conclusion as

CSHO leads to substantially less “noise” in respondents’ attention to price. As our study demonstrates, *ex poste* conditioning for consequentiality and ANA lead to significant improvements in model fit and allowed us to better see our data and the differences in stated preferences between those attending attributes and those ignoring attributes. Our further examination of the ignoring group provided further evidence that the MXL model results used to identify patterns of ANA did an effective job. Model improvements aside, conditioning for ANA ultimately led to higher WTP values, as demonstrated in table 13. The implications are that without accounting for ANA, we would have significantly underestimated WTP for attributes.

Much remains to be learned about how best to identify patterns of and condition data for ANA and the apparent connection between consequentiality, HO and ANA. One potential avenue for further exploration is how CS and HO change the cognitive processes used by respondents when making decisions in a choice experiment. Does increasing a respondent’s perception of consequentiality combined with their promise to reveal truthful answers change their thought processes? Do these methods bring price to the forefront of the thought process or do these methods trigger other types of unexpected choice heuristics? One limitation of this study is that we do not have data on the types of thoughts and decision-making rules being used by respondents. Including such data may further uncover how our *ex ante* approaches are combing with ANA to provide improved models and relatively lower WTP values over the baseline treatment.

References

Bech, M., and Gyrd-Hansen, D. (2005). Effects coding in discrete choice experiments. *Health economics*, *14*(10), 1079-1083.

Bello, M., andAbdulai, A. (2016). Impact of Ex-Ante Hypothetical Bias Mitigation Methods on Attribute Non-Attendance in Choice Experiments. *American Journal of Agricultural Economics*, aav098.

Bliemer, M. C., and Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, *44*(6), 720-734.

Bulte, E., Gerking, S., List, J.A., de Zeeuw, A. (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, 330-342.

Campbell, D., Hutchinson, W.G. and Scarpa, R. (2008). Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental & Resource Economics* 41:401-417.

Caputo, V., Nayga, R.M. and Scarpa, R. (2013). Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics* 57:465-482.

Carlsson, F., Kataria, M. and Lampi, E. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental & Resource Economics* 47:65-89.

Carson, R. T., and T. Groves. “Incentive and Informational Properties of Preference Questions.” *Environmental and Resource Economics* 37(2007):181–210.

Carson, R. T., Groves, T., and List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 171-207.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. “Using Donation Mechanisms to Value Nonuse Benefits from Public Goods.” *Journal of Environmental Economics and Management* 33(1997):151–162.

Choice Metrics (2012) *Ngene 1.1.1 User Manual & Reference Guide*. Sydney, Australia: Choice Metrics Ltd.

Collins, A. T., and Hensher, D. A. (2015). The Influence of Varying Information Load on InferredAttribute Non-Attendance. *Bounded Rational Choice Behaviour: Applications in Transport*, 73.

de-Magistris, T., Gracia, A., and Nayga, R. M. On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments. *American Journal of Agricultural Economics*, *95*(2013) (5), 1136-1154.

Drichoutis, A. C., Vassilopoulos, A., Lusk, J., and Nayga, R. M.. Reference Dependence, Consequentiality and Social Desirability in Value Elicitation: A Study of Fair Labor Labeling. Paper prepared for presentation at the EAAE-AAEA Joint Seminar “Consumer Behavior in a Changing World: Food, Culture, Society” 25-27 March (2015). Naples, Italy.

Erdem, S., Campbell, D., and Hole, A. R. (2015). Accounting for Attribute-Level Non-Attendance in a Health Choice Experiment: Does it Matter? *Health economics*, *24*(7), 773-789.

Harrison, G. W. “Experimental Evidence on Alternative Environmental Valuation Methods.” *Environmental & Resource Economics* 34(2006):125–162.

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

Hensher, D. A., and Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, *30*(2), 133-176.

Hensher, D.A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics* 21:861-878.

Hensher, D.A. (2014). Process heuristics in choice analysis: An editorial. *Journal of Choice Modelling* 11:1-3.

Hensher, D.A. 2008. “Joint Estimation of Process and Outcome in Choice Experiments and Implications for Willingness to Pay.” *Journal of Transport Economics and Policy* 42(2):297-322.

Hensher, D.A. and Greene, W.H. (2010). Non-attendance and dual processing of common metric attributes in choice analysis: A latent class specification. *Empirical Economics* 39:413-426.

Hensher, D.A. and Rose, J.M. (2009). Simplifying choice through attribute preservation or non attendance: Implications for willingness to pay. *Transportation Research Part E* 45:583-590.

Hensher, D.A., Rose, J.M. and Greene, W.H. (2015). Applied choice analysis. Second edition. Cambridge University Press. Cambridge, UK.

Herriges, J., C. Kling, C. C. Liu, and J. Tobias. “What Are the Consequences of Consequentiality?” *Journal of Environmental Economics and Management* 59(2010):67–81.

Hess, S. and Hensher, D.A. (2010). 776 Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B* 44:781-790.

Hess, S., and J. M. Rose. 2008. “Should Reference Alternatives in Pivot Design SC Surveys Be Treated Differently?” *Environmental and Resource Economics* 42(3):297:317

Interis, M., and Petrolia, D. The Effects of Consequentiality in Binary-and Multinomial-Choice Surveys. *Journal of Agricultural and Resource Economics*, *39*(2) (2014), 201-216.

Islam, T., J. J. Louviere, and P. F. Burke (2007). Modeling the Effects of Including/Excluding Attributes in Choice Experiments on Systematic and Random Components. *International Journal of Market Research* 24:289–300.

Jacquemet, N., James, A., Luchini, S., Shogren, J. *Referendum under Oath*. CREQHAM Working Paper 2010-15 (2010).

Jacquemet, N., Joule, R.V., Luchini, S., Shogren, J. *Eliciting Preferences under Oath*. CES Working Paper 09(43) (2009).

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

Joule, R.V., Girandola, F., and Bernard, F. How can people be induced to willingly change their behavior? The path from persuasive communication to binding communication. *Social and Personality Psychology Compass*, *1*(1) (2007), 493-505

Kling, C. L., D. J. Phaneuf, and J. Zhao. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26(2012):3–26.

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

Kulik, J., Carlino, P., The effect of verbal communication and treatment choice on medication compliance in a pediatric setting. *Journal of Behavioral Medicine* 10 (4)(1987), 367e376.

Layton, D.F. and Brown, G. (2000). Heterogeneous preferences regarding global climate change. *The Review of Economics and Statistics* 82:616–24.

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

List, J. A., and C. A. Gallet. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental & Resource Economics* 20(2001):241–254.

Loomis, J., C. Pierce, and M. Manfredo. "Using the Demand for Hunting Licenses to Evaluate Contingent Valuation Estimates of Willingness to Pay." *Applied Economics Letters* 7(2000):435–438.

Lusk, J. L., and T. C. Schroeder. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86(2004):467–482.

Lusk, J. L., Jamal, M., Kurlander, L., Roucan, M., and Taulman, L. A meta-analysis of genetically modified food valuation studies. *Journal of Agricultural and Resource Economics* 30(2005):28-44.

Lusk, J. L., McFadden, B. R., and Rickard, B. J. (2015). Which biotech foods are most acceptable to the public? *Biotechnology journal, 10(2015)*(1), 13-16.

Mitani, Y., and N. E. Flores. “Hypothetical Bias Reconsidered: Payment and Provision Uncertainties in a Threshold Provision Mechanism.” Paper presented at the World Congress on Environmental and Resource Economics, Montreal, Canada, July 1, 2010.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005). “A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation.” *Environmental & Resource Economics* 30(2005):313–325.

Non-GMO Project. www.nongmoproject.org

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.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households’ choices of appliance efficiency level. *The Review of Economics and Statistics* 80:647–57.

Sawtooth, Inc. Sawtooth Software (2015). http://www.sawtoothsoftware.com/. (2015)

Scarpa, R., Campbell, D. and Hutchinson, W.G. (2007). Benefit estimates for landscape improvements: Sequential Bayesian design and respondents' rationality in a choice experiment. *Land Economics* 83:617-634.

Scarpa, R., Gilbride, T.J., Campbell, D. and Hensher, D.A. (2009). Modelling attribute non attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36:151-174.

Scarpa, R., Thiene, M., and Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, *90*(4), 994-1010.

Scarpa, R., Zanoli, R., Bruschi, V. and Naspetti, S. (2013). Inferred and stated attribute nonattendance in food choice experiments. *American Journal of Agricultural Economics* 95:165-180.

Schäfer, J., and K. Strimmer. (2005). A shrinkage approach to large-scale covariance estimation and implications for functional genomics. *Statist. Appl. Genet. Mol. Biol.* 4:32.

Schäfer, J., R. Opgen-Rhein, R., V. Zuber, M. Ahdesmäki, A. P. Duarte Silva, and K. Strimmer. Efficient Estimation of Covariance and (Partial) Correlation. *R Package ‘corpcor’*. Version 1.6.8, 2015. http://strimmerlab.org/software/corpcor/

United States Department of Agriculture (USDA) Economic Research Service (ERS) (2015). Retail prices for beef, pork, and poultry cuts, effs, and dairy products. www.ers.usda.gov

Van Loo, E.J., V. Caputo, R. M. Nayga Jr., and W. Verbeke. 2014. Consumers' valuation of sustainability labels on meat. *Food Policy* 49:137-150.

Van Loo, E.J., Caputo, V., Nayga Jr., R. M., Seo, H.-S., Zhang, B. and Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay and visual attention to attributes. *Ecological Economics* 118:215-225.

Vossler, C. A., and J. Kerkvliet (2003) "A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a Public Referendum." *Journal of Environmental Economics and Management* 45(2003):631–649.

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(2009):338–345.

Vossler, Christian A., and Sharon B. Watson (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior and Organization* 86:137–47.

Vossler, C.A., Doyon, M., Rondeau, D. (2012). Truth in consequences: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4(4), 145-171.

Tables and Figures

Table 1. Choice Experiment Attributes and Levels with Effects Coding

Attributes	**Levels**	**Coding**
Price (4)	$2.99	$2.99
	$6.99	$6.99
	$10.99	$10.99
	$14.99	$14.99
	No-buy	0
GM Content (3)	No information	-1,-1
	Non-GMO verified	1, 0
	Contains GM	0, 1
	No-buy	0, 0
Carbon Footprint (4)	No information	-1,-1,-1
	79 oz CO2e/lb (low)	1, 0, 0
	90 oz CO2e/lb (medium)	0, 1, 0
	112 oz CO2e/lb (high)	0, 0, 1
	No-buy	0, 0, 0
Local (2)	No information	-1
	Local production	1
	No-buy	0

Table 2. Sample Characteristics, Counts and Percentages across the Treatments

	Consequentiality Script (CS)		Honesty Oath (HO)		Combined (CSHO)		Inconsequential Control (IC)		Baseline Control (BC)		Total	
Gender	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
Male	160	31.5%	174	34.5%	158	30.9%	163	32.2%	161	31.9%	816	32.2%
Female	348	68.5%	330	65.5%	354	69.1%	343	67.8%	344	68.1%	1719	67.8%
χ2 = 1.807												
p-value = 0.7712												
Age group	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
18–24 years	27	5.3%	38	7.5%	43	8.4%	36	7.1%	40	7.9%	184	7.3%
25–34 years	121	23.8%	100	19.8%	119	23.2%	110	21.7%	107	21.2%	557	22.0%
35–44 years	94	18.5%	82	16.3%	81	15.8%	90	17.8%	72	14.3%	419	16.5%
45–54 years	76	15.0%	86	17.1%	85	16.6%	87	17.2%	99	19.6%	433	17.1%
55–64 years	88	17.3%	104	20.6%	98	19.1%	86	17.0%	98	19.4%	474	18.7%
65 years or older	102	20.1%	94	18.7%	86	16.8%	97	19.2%	89	17.6%	468	18.5%
χ2 = 17.58												
p-value = 0.615												
Education Level	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
Some Grade School	0	0.0%	0	0.0%	0	0.0%	2	0.4%	0	0.0%	2	0.1%
Some High School	9	1.8%	3	0.6%	5	1.0%	9	1.8%	9	1.8%	35	1.4%
High School Diploma	159	31.3%	170	33.7%	177	34.6%	152	30.0%	157	31.1%	815	32.1%
Associates Degree (2-year degree)	110	21.7%	115	22.8%	93	18.2%	105	20.8%	128	25.3%	551	21.7%
Bachelors Degree (4-year degree)	157	30.9%	140	27.8%	153	29.9%	153	30.2%	126	25.0%	729	28.8%
Masters Degree	58	11.4%	61	12.1%	70	13.7%	64	12.6%	70	13.9%	323	12.7%
Doctoral Degree	15	3.0%	15	3.0%	14	2.7%	21	4.2%	15	3.0%	80	3.2%
χ2 = 29.353												
p-value = 0.2071												

Note: Chi-square test was used to determine if there were differences in sociodemographic 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, and education level.

Table 3. Sample Characteristics, Counts and Percentages across the Treatments, continued

	Consequentiality Script (CS)		Honesty Oath (HO)		Combined (CSHO)		Inconsequential Control (IC)		Baseline Control (BC)		Total	
Income	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
Under $20,000	60	11.8%	63	12.5%	58	11.3%	47	9.3%	62	12.3%	290	11.4%
20,000-39,999	115	22.6%	114	22.6%	124	24.2%	112	22.1%	110	21.8%	575	22.7%
40,000-59,999	107	21.1%	94	18.7%	97	18.9%	106	20.9%	86	17.0%	490	19.3%
60,000-79,999	76	15.0%	79	15.7%	87	17.0%	98	19.4%	87	17.2%	427	16.8%
80,000-99,999	67	13.2%	63	12.5%	57	11.1%	58	11.5%	66	13.1%	311	12.3%
100,000-119,999	30	5.9%	41	8.1%	36	7.0%	31	6.1%	38	7.5%	176	6.9%
120,000-139,999	19	3.7%	15	3.0%	15	2.9%	15	3.0%	21	4.2%	85	3.4%
140,000-159,999	16	3.1%	14	2.8%	11	2.1%	16	3.2%	19	3.8%	76	3.0%
160,000 and above	18	3.5%	21	4.2%	27	5.3%	23	4.5%	16	3.2%	105	4.1%
$\chi2$ = 21.839												
p-value = 0.9116												
Region	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
Northeast	126	24.8%	125	24.8%	126	24.6%	125	24.7%	127	25.1%	629	24.8%
Midwest	126	24.8%	125	24.8%	128	25.0%	125	24.7%	125	24.8%	629	24.8%
South	128	25.2%	128	25.4%	133	26.0%	130	25.7%	126	25.0%	645	25.4%
West	128	25.2%	126	25.0%	125	24.4%	126	24.9%	127	25.1%	632	24.9%
$\chi2$ = 0.255												
p-value = 1.0000												
Race	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
American Indian or Alaska Native	6	1.2%	5	1.0%	6	1.2%	5	1.0%	2	0.4%	24	0.9%
Asian	22	4.3%	21	4.2%	27	5.3%	22	4.3%	19	3.8%	111	4.4%
Black or African American	43	8.5%	43	8.5%	36	7.0%	37	7.3%	44	8.7%	203	8.0%
Native Hawaiian or Other Pacific Islander	0	0.0%	0	0.0%	2	0.4%	0	0.0%	1	0.2%	3	0.1%
White	420	82.7%	423	83.9%	423	82.6%	427	84.4%	428	84.8%	2121	83.7%
Mixed	9	1.8%	7	1.4%	16	3.1%	9	1.8%	7	1.4%	48	1.9%
no response	8	1.6%	5	1.0%	2	0.4%	6	1.2%	4	0.8%	25	1.0%
$\chi2$ = 20.243												
p-value = 0.6829												
Hispanic	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Total	Percent
Hispanic or Latino	40	7.9%	41	8.1%	35	6.8%	39	7.7%	32	6.3%	187	7.4%
Not Hispanic or Latino	468	92.1%	463	91.9%	477	93.2%	467	92.3%	473	93.7%	2348	92.6%
$\chi2$ = 1.708												
p-value = 0.7893												

Note: Chi-square test was used to determine if there were differences in sociodemographic 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 income, region, race, and ethnicity.

Table 4. Distribution of Stated Consequentiality across Treatments

Responses		Consequentiality Script (CS)	Honesty Oath (HO)	Combined (CSHO)	Inconsequential Control (IC)	Baseline Control (BC)	Hypothesis Test	
Consequential ($p>0$)	no.	464	444	460	443	450	χ^2	7.5
	percent	94.7%	93.1%	94.5%	91.7%	94.1%	*p*-value	0.1113
Inconsequential ($p=0$)	no.	26	33	27	40	28	χ^2	103.1
	percent	5.3%	6.9%	5.5%	8.3%	5.9%	*p*-value	***<0.001***
Hypothesis Test	χ^2	23.9	5.5	14.2	64.7	5.3		
	p-value	***<0.001***	***0.0192***	***<0.001***	***<0.001***	***0.0212***		

Note: The null hypothesis of equality in rates of stated consequentiality across treatments is rejected for respondents stating that responses were inconsequential. Hypotheses tests at the bottom of the table are the treatment (cell) by stated consequentiality tests.

Table 5. Mixed Logit (MXL) Baseline Models across Five Treatments

		Consequentiality Script (CS)			Honesty Oath (HO)			Combined (CSHO)			Inconsequential Control (IC)			Baseline Control (BC)		
Variables	Coeff.	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values
PRICE	μ	-0.48 ***	0.02	0.000	-0.47 ***	0.03	0.000	-0.54 ***	0.03	0.000	-0.49 ***	0.03	0.000	-0.45 ***	0.03	0.000
	σ	0.35 ***	0.03	0.000	0.40 ***	0.03	0.000	0.40 ***	0.03	0.000	0.35 ***	0.03	0.000	0.37 ***	0.03	0.000
NON-GM (NGE)	μ	1.10 ***	0.21	0.000	1.18 ***	0.19	0.000	1.16 ***	0.21	0.000	1.03 ***	0.19	0.000	1.14 ***	0.17	0.000
	σ	2.60 ***	0.18	0.000	2.37 ***	0.20	0.000	2.48 ***	0.17	0.000	2.27 ***	0.17	0.000	2.21 ***	0.18	0.000
GM (GME)	μ	-0.93 ***	0.13	0.000	-0.94 ***	0.12	0.000	-0.95 ***	0.14	0.000	-0.79 ***	0.12	0.000	-0.93 ***	0.12	0.000
	σ	1.56 ***	0.13	0.000	1.46 ***	0.14	0.000	1.57 ***	0.14	0.000	1.42 ***	0.14	0.000	1.24 ***	0.11	0.000
LOWCO2 (LOE)	μ	0.22 **	0.10	0.034	0.15	0.11	0.164	0.28 ***	0.10	0.008	0.05	0.11	0.636	0.08	0.10	0.448
	σ	0.56 ***	0.15	0.000	0.53 ***	0.15	0.001	0.72 ***	0.17	0.000	0.51 ***	0.14	0.000	0.69 ***	0.16	0.000
MEDIUMCO2 (MDE)	μ	-0.01	0.10	0.885	-0.02	0.09	0.844	-0.12 ***	0.10	0.235	-0.02	0.10	0.857	0.01	0.10	0.889
	σ	0.42 **	0.17	0.014	0.14	0.17	0.419	0.24	0.19	0.217	0.23	0.14	0.101	0.17	0.16	0.303
HIGHCO2 (HIE)	μ	-0.06	0.09	0.510	-0.04	0.10	0.650	-0.06	0.09	0.518	0.08	0.09	0.337	0.08	0.08	0.352
	σ	0.70 ***	0.16	0.000	0.74 ***	0.23	0.002	0.60 **	0.24	0.012	0.55 ***	0.15	0.000	0.72 ***	0.14	0.000
LOCAL (LCE)	μ	0.24 ***	0.06	0.000	0.29 ***	0.06	0.000	0.28 ***	0.07	0.000	0.15 **	0.06	0.016	0.29 ***	0.06	0.000
	σ	0.53 **	0.27	0.048	0.41 ***	0.14	0.004	0.59 ***	0.15	0.000	0.46 ***	0.09	0.000	0.37	0.29	0.205
No-buy (NONE)		-4.88 ***	0.27	0.000	-5.67 ***	0.33	0.000	-5.89 ***	0.34	0.000	-5.18 ***	0.30	0.000	-4.82 ***	0.29	0.000
Error Component	σ	3.00 ***	0.26	0.000	3.56 ***	0.31	0.000	3.48 ***	0.26	0.000	3.47 ***	0.28	0.000	3.00 ***	0.26	0.000
N. parameters			37			37			37			37			37	
Log likelihood			-2782.95			-2711.13			-2693.75			-2760.59			-2771.45	
BIC			5872.03			5727.39			5693.40			5826.78			5848.11	
BIC/N			1.50			1.50			1.46			1.51			1.53	
AIC			5639.90			5496.25			5461.49			5595.18			5616.90	
AIC/N			1.44			1.44			1.40			1.45			1.47	
AIC3			5676.90			5533.25			5498.49			5653.90			5653.90	
AIC3/N			1.45			1.45			1.41			1.48			1.48	

***, **, * Significance at 1%, 5%, 10% level

Table 6. Mixed Logit (MXL) Models Conditioned for Consequentiality across Five Treatments

Variables	Coeff.	Consequentiality Script (CS)			Honesty Oath (HO)			Combined (CSHO)			Inconsequential Control (IC)			Baseline Control (BC)		
		Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values
PRICE	μ	-0.46 ***	0.02	0.000	-0.47 ***	0.03	0.000	-0.57 ***	0.03	0.000	-0.47 ***	0.03	0.000	-0.43 ***	0.03	0.000
	σ	0.34 ***	0.03	0.000	0.41 ***	0.03	0.000	0.42 ***	0.03	0.000	0.35 ***	0.03	0.000	0.37 ***	0.03	0.000
NON-GM (NGE)	μ	1.10 ***	0.21	0.000	1.24 ***	0.19	0.000	1.21 ***	0.22	0.000	0.97 ***	0.19	0.000	1.19 ***	0.17	0.000
	σ	2.60 ***	0.18	0.000	2.33 ***	0.20	0.000	2.65 ***	0.19	0.000	2.15 ***	0.19	0.000	2.10 ***	0.19	0.000
GM (GME)	μ	-0.93 ***	0.13	0.000	-0.95 ***	0.13	0.000	-0.96 ***	0.15	0.000	-0.73 ***	0.12	0.000	-0.92 ***	0.12	0.000
	σ	1.58 ***	0.13	0.000	1.48 ***	0.15	0.000	1.66 ***	0.15	0.000	1.36 ***	0.15	0.000	1.20 ***	0.11	0.000
LOWCO2 (LOE)	μ	0.16	0.11	0.133	0.12	0.11	0.259	0.33 ***	0.11	0.004	0.05	0.11	0.637	0.11	0.10	0.284
	σ	0.60 ***	0.17	0.000	0.51 ***	0.16	0.001	0.82 ***	0.16	0.000	0.48 **	0.20	0.017	0.65 ***	0.17	0.000
MEDIUMCO2 (MDE)	μ	0.00	0.10	0.971	-0.03	0.09	0.710	-0.10 ***	0.10	0.318	0.00	0.10	0.969	0.04	0.10	0.692
	σ	0.39 ***	0.13	0.003	0.14	0.18	0.431	0.25 *	0.14	0.080	0.23	0.20	0.251	0.16	0.18	0.391
HIGHCO2 (HIE)	μ	-0.02	0.09	0.864	-0.01	0.10	0.900	-0.12	0.10	0.226	0.07	0.09	0.459	0.03	0.09	0.687
	σ	0.70 ***	0.19	0.000	0.76 ***	0.20	0.000	0.67 ***	0.14	0.000	0.53 *	0.28	0.056	0.70 ***	0.23	0.003
LOCAL (LCE)	μ	0.22 ***	0.06	0.000	0.29 ***	0.07	0.000	0.30 ***	0.07	0.000	0.18 ***	0.06	0.005	0.29 ***	0.06	0.000
	σ	0.49 *	0.25	0.053	0.39 **	0.19	0.036	0.65 ***	0.13	0.000	0.48 **	0.20	0.016	0.35	0.29	0.216
No-buy (NONE)		-4.86 ***	0.28	0.000	-5.67 ***	0.35	0.000	-6.01 ***	0.36	0.000	-5.12 ***	0.31	0.000	-4.85 ***	0.30	0.000
Error Component	σ	2.73 ***	0.26	0.000	3.63 ***	0.33	0.000	3.62 ***	0.30	0.000	3.50 ***	0.30	0.000	3.10 ***	0.28	0.000
N. parameters		37			37			37			37			37		
Log likelihood		-2654.75			-2533.04			-2523.94			-2551.31			-2635.17		
BIC		5613.61			5368.56			5351.67			5405.02			5573.32		
BIC/N		1.51			1.51			1.45			1.53			1.55		
AIC		5383.50			5140.07			5121.88			5176.62			5344.34		
AIC/N		1.45			1.45			1.39			1.46			1.48		
AIC3		5420.50			5177.07			5158.88			5381.34			5381.34		
AIC3/N		1.46			1.46			1.40			1.49			1.49		

***, **, * Significance at 1%, 5%, 10% level

Table 7. Distribution of Inferred Attribute Non-Attendance across Treatments

Attributes		Consequentiality Script (CS)	Honesty Oath (HO)	Combined (CSHO)	Inconsequential Control (IC)	Baseline Control (BC)	Hypothesis Test	
Price	no.	33	53	35	35	45	χ^2	9.1
	percent	7.1%	11.9%	7.6%	7.9%	10.0%	*p*-value	0.0598
GM Content	no.	168	192	181	225	197	χ^2	22.2
	percent	36.2%	43.2%	39.4%	50.8%	43.8%	*p*-value	***<0.001***
Carbon Footprint	no.	345	394	368	330	368	χ^2	39.4
	percent	74.4%	88.7%	80.0%	74.5%	81.8%	*p*-value	***<0.001***
Local	no.	235	151	175	237	122	χ^2	93.1
	percent	50.7%	34.0%	38.0%	53.5%	27.1%	*p*-value	***<0.001***
Hypothesis Test	χ^2	1.9	4.6	1.0	0.6	0.7		
(Price)	*p*-value	0.1645	***0.0312***	0.3232	0.4566	0.4164		
Hypothesis Test	χ^2	7.9	0.1	2.0	12.0	0.3		
(GM Content)	*p*-value	***0.0051***	0.7814	0.1579	***0.0005***	0.2583		
Hypothesis Test	χ^2	8.1	25.1	0.0	7.4	1.1		
(Carbon Footprint)	*p*-value	***0.0043***	***<0.001***	0.9283	***0.0065***	0.2976		
Hypothesis Test	χ^2	18.7	8.4	1.3	29.5	36.2		
(Local)	*p*-value	***<0.001***	***0.0038***	0.2462	***<0.001***	***<0.001***		

Note: The null hypothesis of equality in rates of ANA inferred across treatments is rejected for all attributes except price. Hypotheses tests at the bottom of the table are the treatment (cell) by attribute tests.

Table 8. Mixed Logit (MXL) Models Conditioned for Consequentiality and ANA across Five Treatments, Attending Group

		Consequentiality Script (CS)			Honesty Oath (HO)			Combined (CSHO)			Inconsequential Control (IC)			Baseline Control (BC)		
Variables	Coeff.	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values	Estimate	Standard Errors	*p*-values
PRICE	μ	-0.54 ***	0.02	0.000	-0.58 ***	0.03	0.000	-0.63 ***	0.03	0.000	-0.56 ***	0.03	0.000	-0.52 ***	0.03	0.000
	σ	0.29 ***	0.02	0.000	0.34 ***	0.03	0.000	0.31 ***	0.02	0.000	0.29 ***	0.02	0.000	0.34 ***	0.03	0.000
NON-GM (NGE)	μ	2.99 ***	0.27	0.000	3.23 ***	0.26	0.000	3.21 ***	0.33	0.000	3.43 ***	0.24	0.000	3.12 ***	0.19	0.000
	σ	3.54 ***	0.26	0.000	2.79 ***	0.24	0.000	4.37 ***	0.32	0.000	2.53 ***	0.23	0.000	1.96 ***	0.19	0.000
GM (GME)	μ	-1.99 ***	0.18	0.000	-2.47 ***	0.21	0.000	-2.38 ***	0.24	0.000	-2.34 ***	0.20	0.000	-2.12 ***	0.16	0.000
	σ	2.27 ***	0.22	0.000	2.12 ***	0.22	0.000	2.91 ***	0.25	0.000	1.86 ***	0.17	0.000	1.41 ***	0.15	0.000
LOWCO2 (LOE)	μ	0.86 ***	0.28	0.002	1.50 **	0.58	0.010	1.27 ***	0.39	0.001	0.29	0.38	0.449	0.56 **	0.28	0.049
	σ	1.52 ***	0.30	0.000	1.53 *	0.79	0.052	2.29 ***	0.47	0.000	1.43 **	0.62	0.021	1.17 ***	0.37	0.002
MEDIUMCO2 (MDE)	μ	-0.15	0.35	0.666	-0.04	0.42	0.931	-0.49 ***	0.31	0.118	-0.31	0.26	0.239	-0.04	0.28	0.900
	σ	1.49 ***	0.34	0.000	0.82	0.92	0.373	1.03 *	0.52	0.050	0.46	0.52	0.379	0.66	0.61	0.275
HIGHCO2 (HIE)	μ	-0.34	0.31	0.265	-2.32 **	1.03	0.025	-0.93 **	0.42	0.028	0.08	0.41	0.840	-0.28	0.32	0.379
	σ	1.42 ***	0.33	0.000	2.79 **	1.40	0.046	3.06 ***	0.49	0.000	1.56 **	0.79	0.048	1.17 ***	0.35	0.001
LOCAL (LCE)	μ	0.85 ***	0.09	0.000	0.67 ***	0.08	0.000	0.76 ***	0.11	0.000	0.57 ***	0.10	0.000	0.64 ***	0.07	0.000
	σ	0.70 **	0.31	0.024	0.30	0.24	0.211	1.02 ***	0.26	0.000	0.76	0.53	0.156	0.19	0.14	0.171
No-buy (NONE)		-5.12 ***	0.23	0.000	-5.76 ***	0.28	0.000	-5.84 ***	0.27	0.000	-5.43 ***	0.25	0.000	-4.96 ***	0.24	0.000
Error Component	σ	2.51 ***	0.21	0.000	3.29 ***	0.25	0.000	3.05 ***	0.21	0.000	3.36 ***	0.25	0.000	2.87 ***	0.25	0.000
N. parameters		37			37			37			37			37		
Log likelihood		-2407.28			-2214.20			-2241.74			-2265.50			-2360.08		
BIC		5118.68			4730.89			4787.27			4833.41			5023.14		
BIC/N		1.38			1.33			1.30			1.36			1.40		
AIC		4888.57			4502.40			4557.47			4605.00			4794.16		
AIC/N		1.32			1.27			1.24			1.30			1.33		
AIC3		4925.57			4539.40			4594.47			4831.16			4831.16		
AIC3/N		1.33			1.28			1.25			1.34			1.34		

***, **, * Significance at 1%, 5%, 10% level

Table 9. Mixed Logit (MXL) Models Conditioned for Consequentiality and ANA across Five Treatments, Ignoring Group

		Consequentiality Script (CS)			Honesty Oath (HO)			Combined (CSHO)			Inconsequential Control (IC)			Baseline Control (BC)		
Variables	Coeff.	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values	Estimate	Standard Errors	*p* -values
PRICE	μ	0.34 ***	0.04	0.000	0.31 ***	0.04	0.000	0.32 ***	0.07	0.000	0.31 ***	0.05	0.000	0.33 ***	0.04	0.000
	σ	0.02	0.15	0.915	0.08 **	0.04	0.030	0.09	0.09	0.319	0.06	0.08	0.409	0.02	0.15	0.878
NON-GM (NGE)	μ	-1.15 ***	0.14	0.000	-1.38 ***	0.16	0.000	-1.29 ***	0.18	0.000	-1.37 ***	0.13	0.000	-1.59 ***	0.16	0.000
	σ	0.41 *	0.24	0.087	0.77	1.40	0.584	0.49	1.54	0.749	0.82 ***	0.20	0.000	0.48	1.40	0.730
GM (GME)	μ	0.45 ***	0.12	0.000	0.67 ***	0.10	0.000	0.59 ***	0.14	0.000	0.78 ***	0.10	0.000	0.72 ***	0.12	0.000
	σ	0.40	0.74	0.587	0.18	1.47	0.904	0.31	1.69	0.855	0.20	0.27	0.463	0.37 *	0.22	0.085
LOWCO2 (LOE)	μ	-0.32 ***	0.09	0.001	-0.34 ***	0.08	0.000	-0.32 ***	0.09	0.001	-0.39 ***	0.08	0.000	-0.33 ***	0.09	0.000
	σ	0.16	0.85	0.850	0.02	1.43	0.986	0.18	1.30	0.891	0.26	0.58	0.658	0.17	0.87	0.847
MEDIUMCO2 (MDE)	μ	-0.16 *	0.09	0.077	-0.16 **	0.08	0.050	-0.16 ***	0.09	0.093	-0.13 *	0.08	0.093	-0.11	0.09	0.215
	σ	0.24	0.76	0.753	0.31	1.02	0.760	0.29	1.22	0.810	0.13	0.77	0.866	0.20	0.24	0.388
HIGHCO2 (HIE)	μ	0.30 ***	0.09	0.001	0.32 ***	0.09	0.000	0.29 ***	0.10	0.003	0.24 ***	0.08	0.003	0.26 ***	0.08	0.001
	σ	0.25	0.78	0.748	0.36	0.85	0.672	0.28	0.95	0.770	0.16	0.25	0.528	0.32	0.22	0.150
LOCAL (LCE)	μ	-0.37 ***	0.07	0.000	-0.63 ***	0.09	0.000	-0.50 ***	0.08	0.000	-0.33 ***	0.07	0.000	-0.59 ***	0.11	0.000
	σ	0.31	0.68	0.643	0.36	1.18	0.761	0.26	0.96	0.791	0.36	0.33	0.272	0.29	0.26	0.261
No-buy (NONE)		-0.56 ***	0.13	0.000	-0.58 ***	0.13	0.000	-0.66 ***	0.12	0.000	-0.53 ***	0.14	0.000	-0.42 ***	0.12	0.001
Error Component	σ	1.84 ***	0.16	0.000	2.01 ***	0.18	0.000	1.94 ***	0.18	0.000	2.01 ***	0.18	0.000	2.00 ***	0.17	0.000
N. parameters		37			37			37			37			37		
Log likelihood		-3403.23			-3115.00			-3325.45			-3183.83			-3175.55		
BIC		7110.57			6532.49			6954.69			6670.07			6654.08		
BIC/N		1.92			1.84			1.89			1.88			1.85		
AIC		6880.46			6304.00			6724.90			6441.67			6425.09		
AIC/N		1.85			1.77			1.83			1.82			1.78		
AIC3		6917.46			6341.00			6761.90			6462.09			6462.09		
AIC3/N		1.86			1.79			1.84			1.80			1.80		

***, **, * Significance at 1%, 5%, 10% level

Table 10. WTP Space Model Estimates of Marginal WTP Estimates ($/lb for Chicken)

Treatment	Coefficient		Standard Errors	p-value	Differences in WTP Estimates relative to: Baseline Control (BC)	Inconsequential Control (IC)
Consequentiality Script (CS)						
NoGM	4.66	***	0.58	0.000	-2.35	-1.63
GM	-3.43	***	0.40	0.000	-1.49	-0.87
LowCO2	2.06	***	0.45	0.000	0.59	0.86
MedCO2	0.13		0.59	0.828	0.06	-0.57
HighCO2	-1.37	**	0.57	0.016	0.46	1.19
Local	1.84	***	0.19	0.000	0.17	0.32
Honesty Oath (HO)						
NoGM	5.13	***	0.49	0.000	-1.88	-1.16
GM	-3.97	***	0.37	0.000	-0.95	-0.33
LowCO2	1.24	*	0.74	0.093	-0.23	0.05
MedCO2	-0.06		0.65	0.930	-0.02	-0.65
HighCO2	-2.35	**	1.05	0.026	1.44	2.17
Local	1.50	***	0.12	0.000	-0.17	-0.02
Combined (CS + HO)						
NoGM	3.62	***	0.59	0.000	-3.39	-2.67
GM	-2.89	***	0.39	0.000	-2.03	-1.41
LowCO2	0.93	*	0.54	0.083	-0.54	-0.26
MedCO2	-0.51		0.46	0.268	0.44	-0.19
HighCO2	-0.72		0.64	0.259	-0.18	0.55
Local	1.27	***	0.18	0.000	-0.40	-0.26
Inconsequential Control (IC)						
NoGM	6.29	***	0.45	0.000	-0.72	n/a
GM	-4.30	***	0.37	0.000	-0.62	n/a
LowCO2	1.20	**	0.53	0.023	-0.28	n/a
MedCO2	-0.70	**	0.34	0.041	0.63	n/a
HighCO2	-0.17		0.50	0.731	-0.73	n/a
Local	1.52	***	0.21	0.000	-0.15	n/a
Baseline Control (BC)						
NoGM	7.01	***	0.37	0.000	n/a	0.72
GM	-4.92	***	0.34	0.000	n/a	0.62
LowCO2	1.47	***	0.54	0.006	n/a	0.28
MedCO2	-0.07		0.51	0.884	n/a	-0.63
HighCO2	-0.90	*	0.50	0.069	n/a	0.73
Local	1.67	***	0.12	0.000	n/a	0.15

Note: all parameter distibutions had standard deviations significant at the 1% level with the exception of Local in the Baseline Control which was significant at the 5% level

Table 11. Hypotheses Tests in WTP Space ($/lb for Boneless Skinless Chicken Breast)

Hypotheses Tests	Coefficient[b]		Standard Error	p-value
$H0_1$[a] ($WTP^{CS} - WTP^{BC}$) = 0				
nogm x dtreat$_{CS}$	-0.74	***	0.20	0.000
gm x dtreat$_{CS}$	0.50	***	0.15	0.001
lowCO2 x dtreat$_{CS}$	-0.07		0.21	0.742
medCO2 x dtreat$_{CS}$	0.27		0.23	0.237
highCO2 x dtreat$_{CS}$	-0.05		0.22	0.820
local x dtreat$_{CS}$	0.03		0.06	0.645
$H0_2$[a] ($WTP^{HO} - WTP^{BC}$) = 0				
nogm x dtreat$_{HO}$	-0.48	***	0.17	0.006
gm x dtreat$_{HO}$	0.26	*	0.14	0.070
lowCO2 x dtreat$_{HO}$	-0.17		0.26	0.513
medCO2 x dtreat$_{HO}$	-0.13		0.24	0.597
highCO2 x dtreat$_{HO}$	-0.15		0.27	0.578
local x dtreat$_{HO}$	-0.04		0.05	0.423
$H0_3$ ($WTP^{CSHO} - WTP^{BC}$) = 0				
nogm x dtreat$_{CSHO}$	-0.96	***	0.21	0.000
gm x dtreat$_{CSHO}$	0.61	***	0.16	0.000
lowCO2 x dtreat$_{CSHO}$	-0.22		0.23	0.338
medCO2 x dtreat$_{CSHO}$	-0.31		0.21	0.134
highCO2 x dtreat$_{CSHO}$	0.19		0.24	0.433
local x dtreat$_{CSHO}$	-0.14	**	0.06	0.024
$H0_4$[a] ($WTP^{IC} - WTP^{BC}$) = 0				
nogm x dtreat$_{IC}$	-0.25		0.16	0.116
gm x dtreat$_{IC}$	0.17		0.14	0.222
lowCO2 x dtreat$_{IC}$	-0.20		0.22	0.371
medCO2 x dtreat$_{IC}$	-0.31	*	0.17	0.075
highCO2 x dtreat$_{IC}$	0.39	*	0.20	0.054
local x dtreat$_{IC}$	-0.07		0.06	0.286
$H0_5$[a] ($WTP^{CS} - WTP^{IC}$) = 0				
nogm x dtreat$_{CS}$	-0.42	*	0.23	0.067
gm x dtreat$_{CS}$	0.29	*	0.17	0.084
lowCO2 x dtreat$_{CS}$	0.05		0.22	0.823
medCO2 x dtreat$_{CS}$	0.48	***	0.19	0.010
highCO2 x dtreat$_{CS}$	-0.34		0.23	0.142
local x dtreat$_{CS}$	0.10		0.09	0.233

Note: ∗∗∗, ∗∗, and ∗ denote statistical significance at the 1%, 5%, and 10% levels, respectively.

[a]HO_1, H_{O2}, H_{O3}, H_{O4}, H_{O5}, H_{O6}, H_{O7}, H_{O8}, H_{O9}, and HO_{10} designates the effects of the treatment (dtreat) on the marginal WTP estimate.

[b]Designates the effects of the treatment (dtreat) on the marginal WTP estimate.

Table 12. Hypotheses Tests in WTP Space ($/lb for Boneless Skinless Chicken Breast) continued

Hypotheses Tests	Coefficient[b]		Standard Error	p-value
$H0_6$[a] $(WTP^{HO} - WTP^{IC}) = 0$				
nogm x dtreat$_{HO}$	-0.26		0.20	0.191
gm x dtreat$_{HO}$	0.08		0.16	0.631
lowCO2 x dtreat$_{HO}$	0.05		0.27	0.845
medCO2 x dtreat$_{HO}$	0.18		0.21	0.385
highCO2 x dtreat$_{HO}$	-0.60	**	0.30	0.048
local x dtreat$_{HO}$	0.01		0.07	0.832
$H0_7$[a] $(WTP^{CSHO} - WTP^{IC}) = 0$				
nogm x dtreat$_{CSHO}$	-0.70	***	0.26	0.007
gm x dtreat$_{CSHO}$	0.40	**	0.18	0.028
lowCO2 x dtreat$_{CSHO}$	0.06		0.26	0.828
medCO2 x dtreat$_{CSHO}$	0.02		0.18	0.918
highCO2 x dtreat$_{CSHO}$	-0.29		0.27	0.278
local x dtreat$_{CSHO}$	-0.10		0.09	0.257
$H0_8$[a] $(WTP^{CS} - WTP^{HO}) = 0$				
nogm x dtreat$_{CS}$	-0.20		0.25	0.408
gm x dtreat$_{CS}$	0.22		0.17	0.206
lowCO2 x dtreat$_{CS}$	0.12		0.26	0.635
medCO2 x dtreat$_{CS}$	0.30		0.30	0.321
highCO2 x dtreat$_{CS}$	0.27		0.34	0.424
local x dtreat$_{CS}$	0.10		0.07	0.116
$H0_9$[a] $(WTP^{CS} - WTP^{CSHO}) = 0$				
nogm x dtreat$_{CS}$	0.22		0.29	0.449
gm x dtreat$_{CS}$	-0.09		0.19	0.655
lowCO2 x dtreat$_{CS}$	0.19		0.23	0.404
medCO2 x dtreat$_{CS}$	0.38		0.26	0.143
highCO2 x dtreat$_{CS}$	-0.12		0.29	0.680
local x dtreat$_{CS}$	0.18	**	0.09	0.040
$H0_{10}$[a] $(WTP^{HO} - WTP^{CSHO}) = 0$				
nogm x dtreat$_{HO}$	0.42		0.26	0.110
gm x dtreat$_{HO}$	-0.32	*	0.18	0.083
lowCO2 x dtreat$_{HO}$	-0.01		0.29	0.961
medCO2 x dtreat$_{HO}$	0.19		0.25	0.463
highCO2 x dtreat$_{HO}$	-0.45		0.38	0.238
local x dtreat$_{HO}$	0.07		0.07	0.302

Note: ∗∗∗, ∗∗, and ∗ denote statistical significance at the 1%, 5%, and 10% levels, respectively.
[a]HO_1, H_{O2}, H_{O3}, H_{O4}, H_{O5}, H_{O6}, H_{O7}, H_{O8}, H_{O9}, and HO_{10} designates the effects of the treatment (dtreat) on the marginal WTP estimate.
[b]Designates the effects of the treatment (dtreat) on the marginal WTP estimate.

Table 13. Marginal WTP ($/lb for Boneless Skinless Chicken Breast) Across Treatments and Hypothesis Tests in Preference Space (Fixed Price Coefficient)

Hypotheses Tests	NOGM	GM	LowCO2	MedCO2	HighCO2	Local
$H0_1$ ($WTP^{CS} - WTP^{BC}$) = 0						
[b]WTP^{CS}	5.63	-3.69	1.98	0.34	-1.20	1.65
[c]WTP^{BC}	6.78	-4.50	1.48	0.00	-0.76	1.60
mean difference	1.15	0.82	0.50	0.33	0.43	0.06
p-value[a]	***0.018***	***0.025***	0.236	0.340	0.294	0.383
$H0_2$ ($WTP^{HO} - WTP^{BC}$) = 0						
[d]WTP^{HO}	5.56	-3.98	1.48	-0.53	-1.79	1.38
[c]WTP^{BC}	6.78	-4.50	1.48	0.00	-0.76	1.60
mean difference	1.22	0.53	0.00	0.53	1.03	0.22
p-value[a]	***0.010***	0.110	0.497	0.279	0.214	***0.084***
$H0_3$ ($WTP^{CSHO} - WTP^{BC}$) = 0						
[e]WTP^{CSHO}	5.41	-3.75	1.75	-0.69	-1.19	1.32
[c]WTP^{BC}	6.78	-4.50	1.48	0.00	-0.76	1.60
mean difference	1.37	0.76	0.26	0.69	0.42	0.27
p-value[a]	***0.008***	***0.048***	0.358	0.171	0.302	***0.085***
$H0_4$ ($WTP^{IC} - WTP^{BC}$) = 0						
[f]WTP^{IC}	6.52	-4.50	0.76	-0.60	0.09	1.50
[c]WTP^{BC}	6.78	-4.50	1.48	0.00	-0.76	1.60
mean difference	0.26	0.00	0.72	0.61	0.85	0.10
p-value[a]	0.312	0.496	0.218	0.227	0.163	0.347
$H0_5$ ($WTP^{CS} - WTP^{IC}$) = 0						
[b]WTP^{CS}	5.63	-3.69	1.98	0.34	-1.20	1.65
[f]WTP^{IC}	6.52	-4.50	0.76	-0.60	0.09	1.50
mean difference	0.89	0.81	1.22	0.94	1.28	0.16
p-value[a]	***0.070***	***0.035***	***0.078***	0.121	***0.067***	0.282

[a] *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.

[b]WTP^{CS} indicates mean WTP estimates with Consequentiality Script

[c]WTP^{BC} indicates mean WTP estimates from the baseline control

[d]WTP^{HO} indicates mean WTP estimates with Honesty Priming

[e]WTP^{CSHO} indicates mean WTP estimates with Consequentiality Script and Honesty Oath

[f]WTP^{IC} indicates mean WTP estimates with Inconsequentiality Script (True Hypothetical)

Table 14. Marginal WTP ($/lb for Boneless Skinless Chicken Breast) Across Treatments and Hypothesis Tests in Preference Space (Fixed Price Coefficient)

Hypotheses Tests	NOGM	GM	LowCO2	MedCO2	HighCO2	Local
$H0_6$ ($WTP^{HO} - WTP^{IC}$) = 0						
[d]WTP^{HO}	5.56	-3.98	1.48	-0.53	-1.79	1.38
[f]WTP^{IC}	6.52	-4.50	0.76	-0.60	0.09	1.50
mean difference	0.96	0.53	0.72	0.08	1.88	0.12
p-value[a]	***0.048***	0.130	0.280	0.462	***0.083***	0.313
$H0_7$ ($WTP^{CSHO} - WTP^{IC}$) = 0						
[b]WTP^{CSHO}	5.41	-3.75	1.75	-0.69	-1.19	1.32
[c]WTP^{IC}	6.52	-4.50	0.76	-0.60	0.09	1.50
mean difference	1.11	0.76	0.99	0.08	1.27	0.17
p-value[a]	***0.036***	***0.060***	0.135	0.458	***0.072***	0.259
$H0_8$ ($WTP^{CS} - WTP^{HO}$) = 0						
[b]WTP^{CS}	5.63	-3.69	1.98	0.34	-1.20	1.65
[c]WTP^{HO}	5.56	-3.98	1.48	-0.53	-1.79	1.38
mean difference	0.07	0.29	0.50	0.87	0.60	0.28
p-value[a]	0.450	0.256	0.326	0.166	0.323	***0.080***
$H0_9$ ($WTP^{CS} - WTP^{CSHO}$) = 0						
[b]WTP^{CS}	5.63	-3.69	1.98	0.34	-1.20	1.65
[c]WTP^{CSHO}	5.41	-3.75	1.75	-0.69	-1.19	1.32
mean difference	0.22	0.06	0.23	1.03	0.01	0.33
p-value[a]	0.356	0.454	0.358	***0.076***	0.495	***0.074***
$H0_{10}$ ($WTP^{HO} - WTP^{CSHO}$) = 0						
[d]WTP^{HO}	5.56	-3.98	1.48	-0.53	-1.79	1.38
[e]WTP^{CSHO}	5.41	-3.75	1.75	-0.69	-1.19	1.32
mean difference	0.15	0.23	0.27	0.16	0.60	0.06
p-value[a]	0.396	0.310	0.410	0.420	0.323	0.385

[a] *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.

[b]WTP^{CS} indicates mean WTP estimates with Consequentiality Script

[c]WTP^{BC} indicates mean WTP estimates from the baseline control

[d]WTP^{HO} indicates mean WTP estimates with Honesty Priming

[e]WTP^{CSHO} indicates mean WTP estimates with Consequentiality Script and Honesty Oath

[f]WTP^{IC} indicates mean WTP estimates with Inconsequentiality Script (True Hypothetical)

Table 15. Mean WTP Estimates from Multinomial Logit (MNL) and Mixed Logit (MXL) Models ($/lb for Chicken)

		NoGM	GM	LowCO2	Medium CO2	High CO2	Local
Consequentiality Script (CS)	Base MNL	3.68	-2.50	0.40	0.01	-0.04	0.72
	AA MNL	4.87	-3.26	1.58	0.24	-0.83	1.51
	AA MXL	5.63	-3.69	1.98	0.34	-1.20	1.65
Honesty Oath (HO)	Base MNL	3.38	-2.33	0.19	-0.10	0.16	0.70
	AA MNL	5.12	-3.57	1.50	-0.75	-1.55	1.30
	AA MXL	5.56	-3.98	1.48	-0.53	-1.79	1.38
Combined (CSHO)	Base MNL	3.19	-2.21	0.52	-0.29	0.01	0.60
	AA MNL	4.57	-3.08	1.58	-0.67	-1.10	1.18
	AA MXL	5.41	-3.75	1.75	-0.69	-1.19	1.32
Inconsequential Control (IC)	Base MNL	3.47	-2.06	0.36	-0.34	0.20	0.64
	AA MNL	5.96	-3.91	1.01	-1.03	0.23	1.34
	AA MXL	6.52	-4.50	0.76	-0.60	0.09	1.50
Baseline Control (BC)	Base MNL	3.94	-2.79	0.45	0.04	0.01	0.92
	AA MNL	6.30	-4.22	1.50	-0.13	-0.76	1.56
	AA MXL	6.78	-4.50	1.48	0.00	-0.76	1.60

Figure 1. Example Choice Task

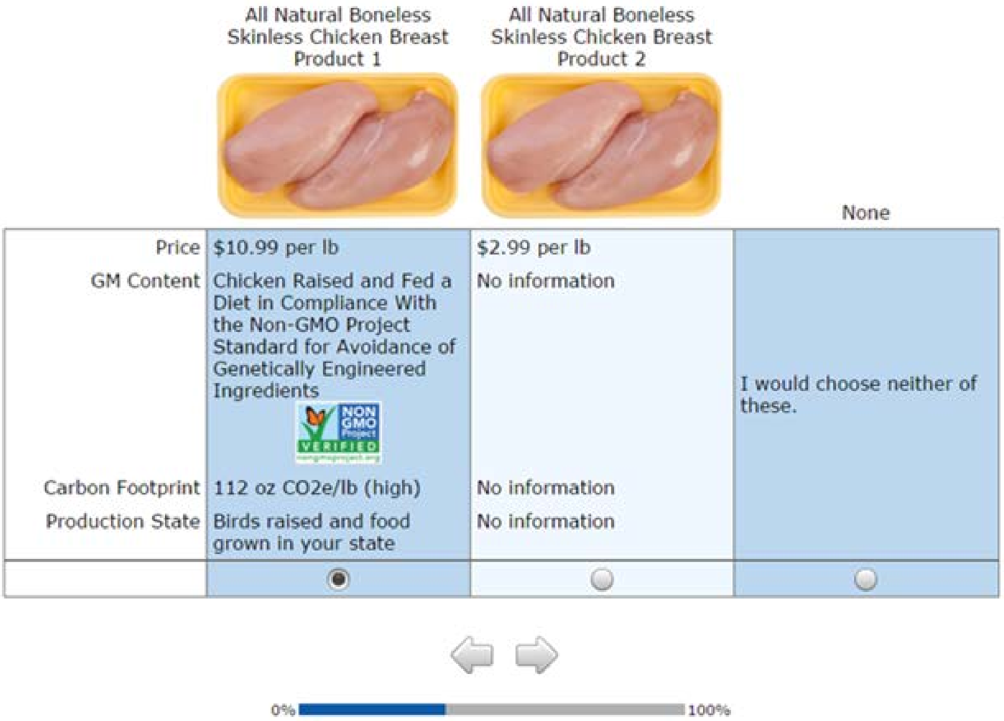

Figure 2. Coefficient of Variation for Conditional Distributions of the Price Attribute for Five Treatments

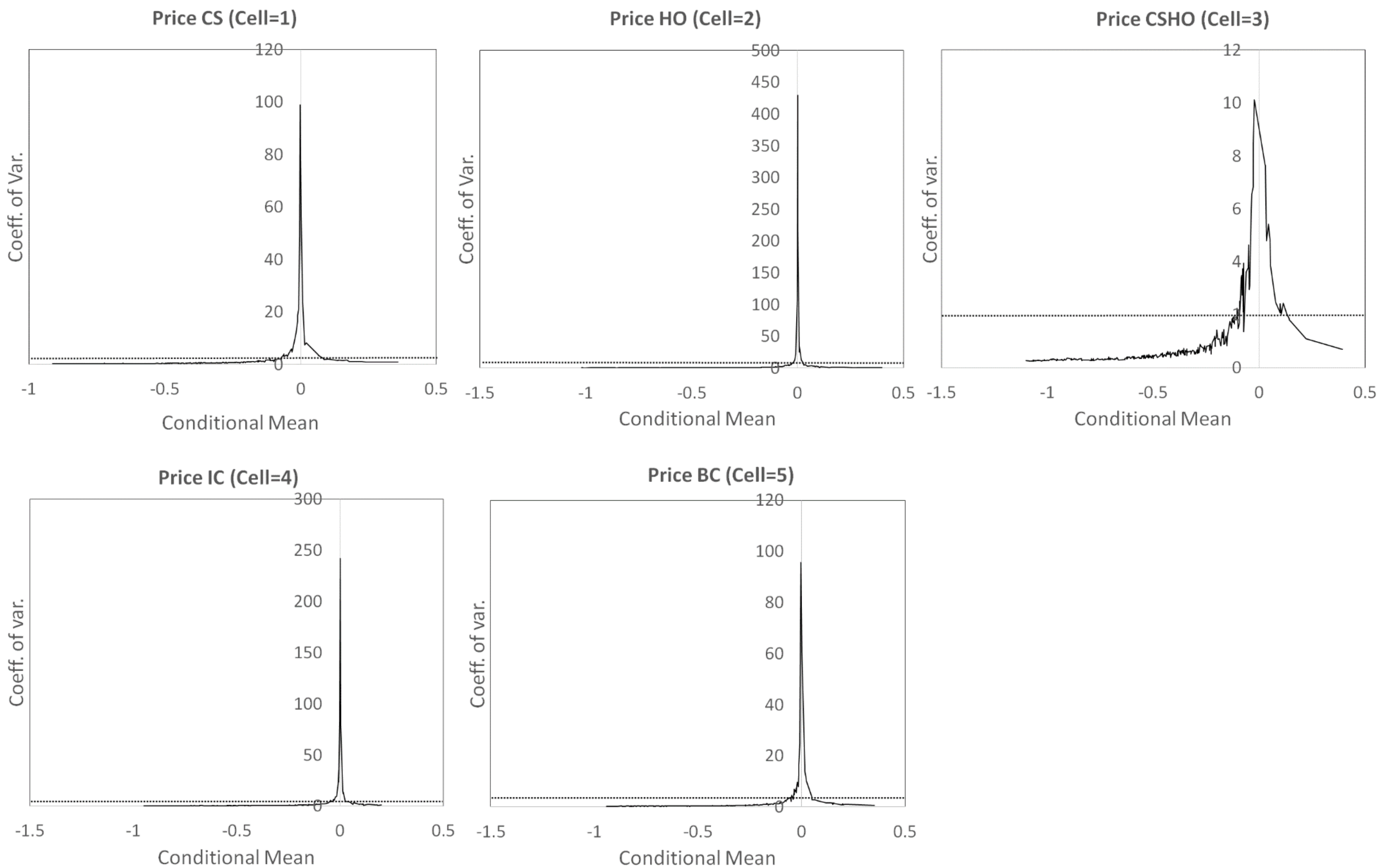

Figure 3. Portion of Respondents Ignoring Attributes in each Treatment