Making the Most of Cheap Talk in an Online Survey

Jerrod Penn, PhD Student, jerrod.penn@uky.edu

Wuyang Hu, Professor, wuyang.hu@uky.edu

Dept. of Agricultural Economics, University of Kentucky

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

Abstract

In stated preference approaches to nonmarket valuation, Cheap Talk is a common method to mitigate Hypothetical Bias, yet the qualities and circumstances that affect its efficacy are still largely undetermined. Online surveys create new opportunity to better engage and test respondents. Prior to a choice experiment, we implement three different Cheap Talk approaches, a standard script, a standard script with an attention-check question, and a script communicated via an online video also with an attention-check question, with the goal of understanding whether the latter augmented cheap talk treatments can improve respondent recognition of hypothetical bias and consequently reduce its magnitude. Results indicate that video cheap talk respondents are more likely to correctly answer the attention-check question compared to the text-based script, indicative of the video's ability to improve attention and comprehension among respondents. …

Introduction

Within environmental economics, stated preference approaches to nonmarket valuation have become commonplace. Carson's work (2011) to document the history and extent of contingent valuation methods (CVM) demonstrates its usefulness. Yet controversies remain (Carson, 2012, Hausman, 2012), specifically the persistent presence of hypothetical bias (HB) which suggests that values elicited from a hypothetical scenario are often greater than an actual scenario. Consequently, mechanisms to allay hypothetical bias have grown in tandem with the use of contingent valuation methods.

Cheap Talk (CT) is one such technique, popularized by Cummings and Taylor (1999), which explicitly explains HB to the respondent and why it might occur as a means of reducing the bias. As an *ex ante*-type approach, CT is shown prior to the valuation component, often imploring the respondent to give an answer as if it were an actual purchase. Examinations since then of CT's effectiveness to reduce

HB is mixed (Loomis, 2014, Silva, et al., 2011). Yet in an effort to guard against HB, many recent studies universally show CT to all respondents across a broad spectrum of products (Food: Gracia, et al. (2014), Lim and Hu (2015); Environment/energy: Kallas and Gil (2015), Herriges, et al. (2010); Health: Gyrd-Hansen (2016)) and locations (China: Guo, et al. (2014), Ortega, et al. (2014); Europe: Van Wezemael, et al. (2014); Zanoli, et al. (2013); Africa: Lagerkvist, et al. (2013); and Oceania: Yao, et al. (2014)).

Usually respondents see CT scripts as a passage of written text just prior to the valuation component. Further, it is taken on faith that the respondent read and understood the CT scripts, while researchers expect at least some respondents will skip such texts in a survey. Others have noted these shortcomings. Lusk (2003) questioned "*Whether consumers will take time to read and assimilate lengthy cheap talk information in the comfort of their own home*." Similarly, Bulte, et al. (2005) said "*Cheap Talk might have little or no effect on some people, like those refusing to read Cheap Talk.*"

Rather than assuming the respondent has read and understood the CT script, we test their comprehension and attentiveness. By immediately questioning the respondent's comprehension, we can apply a kind of ex post correction, evaluating or using the WTP of only those who correctly answered the attention-check question. Online surveys also afford opportunities to overcome weaknesses of a text-based CT, namely the ability to embed video/audio to communicate important information to the respondent. These treatments outline the two main goals of our study, to understand:

1) Does adding an attention check of the respondent's comprehension of the Cheap Talk script strengthen its intended result?
2) Does Cheap Talk spoken aloud as a video improve its effectiveness?

In both cases, efficacy of these CT variants is evaluated by comparing the WTP generated from each of the treatment groups. In the next sections, we outline previous literature on CT followed by the experimental design. Next we present and discuss model results and end with concluding remarks.

Literature Review

The use of online surveys to elicit values has grown rapidly. Gao et al.'s (2015) search of "online survey" or "web-based survey" on Web of Knowledge found that use of online survey data increased from less than 10 to more than 1,000 instances in roughly two decades, and that online surveys of WTP increased from 0 instances in 2001 to 41 by 2012. With its growing use, some have investigated potential differences of online surveys versus other modes. Online surveys can generate additional information on how long a respondent takes to complete each question as well the entire survey. Börger (2015) found that respondents with longer response times have decreased randomness in their choices, while Vista et al. (2009) showed that respondent demographics as well as attitudes and preferences for their studied good were unaffected by response times, either in various parts or the whole survey.

With respect to WTP, internet surveys generate statistically similar values compared to face-to-face interviews (Lindhjem and Navrud, 2011, Marta-Pedroso, et al., 2007, Nielsen, 2011) as well as other, paper-based survey modes (Olsen, 2009, Windle and Rolfe, 2011). Usually, the number of responses collected in a given time frame is many times larger than for a mail or in-person survey. The real-time adjustments to survey presentation are also useful in online surveys. For example, online surveys have display logic that can, for instance, quickly screen out inappropriate respondents, or show customized questions given the respondent's earlier answers. In short, there are real benefits to using online surveys.

The difficulty with online surveys, as in all stated preference methods, is the existence of hypothetical bias, the difference in welfare estimates when comparing hypothetical versus real/actual (when there is a non-zero chance of payment) responses. As previously discussed, multiple investigations have considered the efficacy of CT to eliminate or eliminate HB. As far as we know, no surveys eliciting WTP in an online setting have included a real treatment group, so there is no formal

evidence establishing HB in online surveys nor its magnitude. Yet, evidence of Hypothetical Bias in online surveys still exists. Tonsor and Shupp (2011) show that a significant reduction for WTP of four attributes in their sub-sample of respondents subject to CT.

CT originated from game theory as nonbinding communication between players in a game that does not affect payoffs, but may nonetheless have an effect on players' strategies. It has been implemented in a number of game theory-oriented lab experiment applications. Each of which has to be effective under various game-types and circumstances (Andersson, et al., 2010, Brosig, et al., 2004, Charness and Grosskopf, 2004, Lusk and Hudson, 2004).

Investigations within psychology noted the potential differences in communication within survey design. Tourangeau (1984) noted that "*Oral presentations may be more effective simply because they are more likely to command our attention, but it may be easier to understand written presentations if we bother to read them*." Verbal instructions were found to be a more effective and valuable means of communication relative to pictures (Toepoel and Couper, 2011, Toepoel and Dillman, 2011). Smyth, et al. (2009) found that verbal instructions improved the quality of responses to open-ended questions. The effect of communication mode within game theory and psychology settings gives credence to the potential improvement CT may bring in stated preference valuations.

In Stated Preference approaches, CT is the information provided by the interviewer or written in the survey prior to valuation, which expressly tells the respondent about the existence of hypothetical bias, and to carefully consider their answers as if they were about to make an actual purchase. Early analyses which reminded respondents of substitutes and of their own budget constraints proved ineffective (Loomis 1994, Neill 1995). Many have studied various factors of the elicitation's setting as well as characteristics of CT that may affect its efficacy. See Silva, et al. (2011) for a comprehensive review.

One such vein of research is CT's usefulness at various lengths. Cummings and Taylor (1999)'s original CT script proved effective at 941 words. Per Silva, et al. (2011, Table 1), it also appears to be the most frequently reused (with appropriate modifications) script by other studies. Neither Aadland and Caplan (2003) nor Poe, et al. (2002) found a statistical improvement in WTP from approximately 50-word scripts over the phone. Moreover, Aadland and Caplan (2006) found that a 125-word script lead to significantly *more* HB relative to a shorter 75-word script over the phone. Other short scripts have been effective, with Brummett, et al. (2007) and Silva, et al. (2011) both finding no significant difference between real payment values and hypothetical values that used a 128-word and 211-word CT script, respectively, in field experiment settings. Lastly, Carlsson et al.'s (2005) 113-word mail survey significantly reduced HB for a majority of CE attributes.

Others have focused on the "neutrality" of the CT script, which informs a survey participant that respondents usually *misstate* their values rather than the conventional "overstate" approach in which respondents *overstate* their values compared to real settings (Aadland and Caplan, 2006). For instance, Aadland and Caplan (2006)'s short CT scripts, which exacerbated HB, were both neutral, whereas their overstated script from Aadland and Caplan (2003) effectively reduced HB. Conversely, Silva, et al. (2012) used a neutral CT script that generated WTP statistically equivalent to real WTP. Doyon, et al. (2015) appears to be the only study that used a conventional CT suggesting WTP overstatement that exacerbated WTP relative to the baseline, though not statistically different. The mixture of results suggests additional investigation of neutrality is warranted.

Still others have found that characteristics of the respondent affect CT's ability to reduce or eliminate HB. In List (2001), CT among experienced sportscard dealers was ineffective, but WTP for non-dealers presented with CT was equivalent in the actual and hypothetical groups. Blumenschein, et al. (2008) and Champ, et al. (2009) similarly found that CT was the least effective among well-informed participants. We found one instance of more familiar respondents being more greatly affected by CT

(Ami, et al., 2011)**.** Some works show evidence that only women are affected by CT (Barrage and Lee, 2010, Ladenburg, et al., 2011, Mahieu, 2010).

Finally, some studies show CT only reduce WTP among those offered a high payment level and ineffective at lower payment levels (Brown, et al., 2003, Ladenburg, et al., 2011, Murphy, et al., 2005)

It is important to note that some of the analyses mentioned examined CT without including a real treatment group, respondents who would be asked to make a real payment for the action or option they have chosen in the survey. These can only compare WTP relative to a baseline hypothetical control group without CT. For example, Özdemir, et al. (2009) found that CT WTP was generally lower than the control group for rheumatoid arthritis medication, and Carlsson, et al. (2005)'s Choice Experiment for two food products demonstrated significantly lower WTP for seven of ten attributes compared to the control group.

In these circumstances, the significance CT infers the existence of HB. It may be the extent of HB was is minor enough that leaves little room for CT to 'improve upon' the hypothetical values. Indeed, in their meta-analysis of HB, Penn and Hu (2016) found roughly 25% of all observations without any HB mitigation or correction had a CF (ratio of WTP elicited in a hypothetical situation to WTP elicited in a situation involving real payments) between .81 and 1.2.

More recent studies have continued to test CT, with de-Magistris, et al. (2013), Jacquemet, et al. (2013), and Moser, et al. (2014) all showing its usefulness to reduce HB to varying extents. The range of evidence from the CT literature suggests that CT can, though not guaranteed to, reduce HB, but opportunity exists to improve its chance of success.

Experimental Design and Setting

To fulfill the goals of this study, we test the effect of attention-check and video CT. We used an online survey conducted in May 2015. The survey was built on the Qualtrics platform, who also recruited

the panel of respondents. The survey was advertised as a survey on consumer preferences of hotels, with a particular goal of valuing various environmental practices possible within a hotel.

While there are advantages afforded to online surveys as mentioned earlier, there are also criticisms related to careless or dishonest respondents present in the online panels used to complete such surveys. As per Jones, et al. (2015), our sample of respondents analyzed represent only those passed multiple red herring/trap questions and excluding those who took fewer than 6 minutes to complete the survey, as pretesting of the survey indicated that respondents took on average 15 minutes to complete the survey. On the instruction page of the survey, respondents were informed in exceptionally large, bold-face font that the survey contained audio and video content and were instructed to ensure that those functions were working properly before beginning the survey.[1] The transcript of the Video/paragraph CT appears in the Appendix.

Each respondent of the survey was randomly assigned to one of three treatments:

1. Standard Cheap Talk: baseline group who only saw the CT script as a traditional text-based passage; this is the reference group with which we compare against the other treatments.
2. Attention-Check Cheap Talk: group saw the CT script as well as an attention-check question immediately below the CT script to verify if the respondent understood.
3. Video Cheap Talk: group saw CT as an online video embedded in the survey as well as an attention-check question below video. The video contained all of the same details as the text-based scripts and was about 35 seconds long.

[1] We have no way to guarantee these functions were operational for everyone. Even if the audio was not functional, respondents could still follow the cheap talk script displayed, though in slightly abbreviated version, in the video.

The video and attention-check treatment had the following query, “Most researchers find that in doing studies of hypothetical purchases, respondents typically give values _____ their willingness to pay in an actual purchase.” The first three responses, “Less than,” about equal to,” and “greater than” were displayed in a random order across respondents, while the “I don’t know” option remained in the last position.

At 102 words, the CT script used was equivalent across all three treatments and was arranged as 3 bullets of information. In developing the CT script, we specifically considered previously used *short* scripts including Aadland and Caplan (2006), Brummett, et al. (2007), Carlsson, et al. (2005), Silva, et al. (2011), whose scripts were between 75 words (Aadland and Caplan, 2006) and 211 words (Silva, et al., 2011). This was in part to decrease the text’s block size in order to reduce the likelihood the respondent skipped the script/video and attention-check question altogether. It is worth noting that longer scripts such as Cummings and Taylor (1999) were used in an experimental setting when a moderator has a greater ability to maintain the participant’s attention. The script asked respondents to think about previous prices of hotel rooms they paid for, reminded them that if both were too expensive or unattractive, they could opt-out. In the last bullet, rather than be neutral, the script specifically indicates that respondents WTP in hypothetical questions is usually greater than WTP in a real purchase, which matches the language used in the attention-check question. The third bullet was intentionally presented last in both script and video so that it would be the most recent point made prior to the attention-check question.

Immediately after the CT treatment, respondents completed a choice experiment made up of 8 profiles per respondent. The CE was designed in SAS to optimize D-efficiency for main-effects. The attributes were related to typical and novel aspects of choosing a hotel, with an example choice set appearing in Figure 1. A description of the attributes and their levels are listed in Table 1.

The traditional attributes describing the hotel included the price of the hotel, the average guest review for the hotel, the number of reviews, and whether the hotel included complimentary in-room Wi-Fi or Breakfast. The novel hotel attributes reflected new business practices with social or environmental benefits. These features included the presence of a hotel recycling program, providing full-time employees with a living wage (above the minimum wage), and room cleaning by request. This last feature specified to guests that regular room cleaning services would only occur upon guest request in order to save water and energy. Additionally, rooms could include beneficial indoor plants, either a spider or snake plant. Given that their common names could be poorly interpreted, they were simply described as Plant A and Plant B as well as their scientific names. The final attribute, reports of bed bugs, was actually embedded in a link to "5 recent reviews of each hotel." This reflects reality because only travelers who are interested enough to seek out and read other recent guests' thoughts would have an opportunity to find out about reports of bed bugs in hotels.

To help make the choice sets more comprehensible, 4 of the attributes were communicated pictures and icons; the indoor plants, Wi-Fi, and Breakfast. Except for the bed bug attribute, attributes were described in detail prior to the Choice experiment choice sets, including the corresponding icon for the applicable attributes, and the choice sets themselves appeared in a random order to mitigate order effects (Day, et al., 2012).

Econometric Approach-IN PROGRESS

To model our data of each Cheap Talk treatment, we utilize Mixed Logit Willingness To Pay Space models as introduced in Train and Weeks (2005).

$$U_{njt} = -\lambda_n p_{njt} + (\lambda_n w_n)' x_{jnt} + \varepsilon_{njt}$$

This is a parameterization such that coefficients directly reflect WTP. It is an equivalent expression of the parameter space framework, standard in Choice Experiment modelling. Both WTP and preference space can utilize a mixed logit model (Train, 2009), which allows for controlling for the panel nature of respondents providing multiple answers, does not rely on the IIA assumption, and introduces a distribution for chosen attributes to allow for attribute preference heterogeneity across respondents. The most flexible model includes a distribution in the coefficient of the payment vehicle since respondents have different marginal utilities of income.

A prominent shortcoming in preference space is that if a normal distribution is assumed for the payment vehicle and other attributes, the associated WTP, the ratio of a normally-distributed attribute coefficient to a normally-distributed payment vehicle coefficient, has undefined moments. Instead, assuming a distribution of WTP itself, such that that the ratio is assumed to be normally distributed, rather than the numerator and denominator, avoids these difficulties. WTP Space models are quickly gaining traction in multiple fields of economics (Health: Hole and Kolstad (2012); Food: de-Magistris, et al. (2013); Recreation: Scarpa, et al. (2008); and Transportation: Hensher and Greene (2011)). WTP accounts for (fixed) scale heterogeneity within each model such that comparison of WTP across models is appropriate. Heterogeneity for each attribute's WTP is assumed to follow a normal distribution. At present, we assume no correlation in each attribute's associated WTP.

Preliminary Results

In total, 2069 respondents were included in the comparison of Cheap Talk efficacy, whose summary statistics are reported in Table 2. The average completion time of the survey was approximately twenty minutes.[2]

The sample contains socioeconomic differences relative to the 2014 American Community Survey estimates of the US population (US Census Bureau, 2015). These differences are expected, given that a goal of the sample was to represent US travelers, rather than the entire US population. For example, one requirement for a portion of the sample is to be a business traveler, someone who stayed hotels at least 7 nights for business/work-related reasons, which made up 37.7% of the sample, comparable to industry figures (AHLA, 2016). In our sample, the average respondent is more likely to have acquired higher levels of education, has a higher household income, is older than average, and is more likely to be male. Business and leisure travelers have higher than average household incomes, at $127,000 and $99,000, respectively. In fact, from the entire sample 13.6% of respondents reported staying in hotels for business/work-related reasons at least 25 nights per year.

Examining each treatment, the control group (without any CT) had a sample size of 687, text-based attention-check group was 735, and the remaining 646 were video respondents. In the text-based group, 49.5% (364) correctly answered the attention-check question on the usual direction of HB. In the video group, 77.6% (501) correctly answered the question. This gives at least initial credibility to the usefulness of videos in online surveys to promote online panel respondents' engagement and comprehension of presented survey content and instructions.

While apparent differences exist between the sample and the US population, the various treatment groups appear comparable based on their socioeconomic characteristics. Based on a p-value<.05, almost no differences emerge between the treatment groups. The only significant difference

[2] This measure ignores respondents who completed the survey in more than two hours, about 1% of the sample, who we suspect completed the survey in multiple sessions.

is a comparison of all respondents in the text attention-check treatment group, and the subsect of text attention-check that correctly answered the attention-check question. Significantly fewer respondents with a high school education correctly answered the question. There was also marginally significant evidence that a higher proportion of respondents with advanced degrees correctly answered the attention-check. This gives some mild evidence that more educated individuals will correctly answer in the text-comprehension treatment. A similar, but insignificant pattern follows in the video cheap talk. The similarity of the observable characteristic across treatment groups provides some assurance that differences in WTP across groups is due to the differences in the treatments themselves.

Preliminary Model Results

Results for each treatment's mixed logit WTP Space model appears in Table 2. In addition to the models based on all text-based attention-check respondents and all video respondents, we also include results based on the sub-sample of correct respondents in each group. Virtually all attributes are statistically significant to respondents and follow standard patterns. For example, the negative WTP for the opt-out constant indicates lost value from not being able to stay in a hotel room. Further, respondents place more value on a hotel that has an excellent (5 out of 5) average guest reviews relative to good (4 out of 5) or average (3 out of 5) guest reviews.

Turning to the estimated WTP per attribute across models, we see largely similar values across each treatment and sub-treatment. Giving the correct response does not seem to have a consistent, discernable effect on WTP for either the video or text-based treatments. Even though the percentage of respondents who correctly answered the attention-check question is much higher in the video versus text-based cheap talk, WTP is remarkably similar for the two groups. In the case of attention-check and video treatments relative to their subsamples of correct respondents only, we again see largely similar results.

To test for significant differences, we pooled the Standard CT treatment twice, once with each the Video CT treatment group, and again with the Attention-Check CT data. In each case, we generate interaction terms of the attributes with the respective CT treatment. In these separate models, these dummies are not assumed to have a distribution, and instead remain fixed. As WTP results from Table 3, WTP for attributes in Standard CT are generally larger than the others. A statistically significant and negative interaction term indicates that the particular CT treatment significantly affected WTP for that attribute. Model results from the two pooled datasets are displayed in Table 4.

Based on the pooled model of Standard CT and Video CT, we see a mixture of both positive and negative signs on WTP coefficients, but almost all are not significant. The one exception is the Opt-Out Constant, with a p-value=.051, in which WTP among video respondents decreased an additional $22. The video CT may make respondents more sensitive to the combinations of attributes presented in a particular scenario, affecting the likelihood of deciding to opt-out of either hotel offered. Similar results persist in the pooled standard CT with attention-check CT. Only two interaction terms showed evidence of significantly affecting WTP, Excellent and reports of Bed Bugs. In both cases, the attention-check interaction decreased WTP as one would expect when presuming overstated WTP.

As a supplement, we used an alternative approach to test for differences in WTP across samples is to use parameter space model results to formulate Krinsky-Robb confidence intervals, which can be used as the basis for a complete combinatorial test (Poe, et al., 2005). These results appear in the final column of Table 4. First, we inspect the results of the control group, those who only received the cheap talk as a block of text relative to other treatment groups. For every pairwise comparison, WTP in the standard CT group is either similar and not significantly different, or significantly larger than other treatment groups' WTP. We interpret this as evidence of Hypothetical Bias.

Preliminary Concluding Remarks

It is important to note that because we did not include a treatment group who faced actual, binding choices, we cannot establish the magnitude of Hypothetical Bias. Furthermore, the baseline treatment group also included a Cheap Talk such that there is no way to establish how much the standard text-based CT script may have lowered WTP relative to a true baseline hypothetical treatment group without any CT. Nevertheless, because our goal is to examine potential opportunities to improve CT efficacy, several results are apparent:

Based on the reduction in the WTP per attribute in the attention-check and video treatment groups, there is evidence of HB present in the baseline treatment that included a standard CT script.

Even while the percentage of respondents who correctly answered the attention-check is much higher in the video relative to text-based treatment, both versions have similar WTP results. This means that using a text-based attention-check CT can still improve survey validity and may be a useful addition even in survey modes that cannot readily incorporate a video such mail or in-person intercepts.

Since it is possible that respondents could have skipped, another future improvement would be to measure how long before the respondent clicks to advance past the CT. For example, in text-based approaches, a shorter average time could reflect a greater proportion of those who skip reading the text. It is also possible to include an automated timer which prohibits the respondent from continuing the survey until a certain amount of time has passed, effectively 'forcing' respondents to watch the video. Conversely, by allowing respondents to skip CT in all three treatments, and still observing substantive reductions in WTP, supports the value of using more cognitively-engaging methods to instruct survey respondents.

Appendix: Cheap Talk Treatment with Attention Check

As you answer the next few questions, please keep in mind three things.

- First, consider what you normally spend on a hotel room. On a typical leisure trip, what price would you usually pay for a hotel room per night?
- Second, it's perfectly fine if you are not willing to pay for either hotel if both are too expensive, just 'Choose Neither.'
- And third, previous surveys show that the amount people say they are willing to pay is greater than the amount of a real purchase. As you go through the following scenarios, please imagine you are actually booking a hotel room.

Most researchers find that in doing studies of hypothetical purchases, respondents typically give values _____ their willingness to pay in an actual purchase.

- o About equal to
- o Less than
- o Greater than
- o I don't know

Cited Literature

Aadland, D., and A.J. Caplan. 2006. "Cheap talk reconsidered: New evidence from CVM." *Journal of Economic Behavior & Organization* 60(4):562-578.

---. 2003. "Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias." *American Journal of Agricultural Economics* 85(2):492-502.

American Hotel & Lodging Association (2016) Lodging Industry Trends 2015.

Ami, D., F. Aprahamian, O. Chanel, and S. Luchini. 2011. "A test of cheap talk in different hypothetical contexts: The case of air pollution." *Environmental and Resource Economics* 50(1):111-130.

Andersson, O., M.M. Galizzi, T. Hoppe, S. Kranz, K. Van Der Wiel, and E. Wengström. 2010. "Persuasion in experimental ultimatum games." *Economics letters* 108(1):16-18.

Barrage, L., and M.S. Lee. 2010. "A penny for your thoughts: Inducing truth-telling in stated preference elicitation." *Economics letters* 106(2):140-142.

Blumenschein, K., G.C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. 2008. "Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment*." *The Economic Journal* 118(525):114-137.

Börger, T. 2015. "Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment." *Environmental and Resource Economics*:1-25.

Brosig, J., J. Weimann, and C.-L. Yang. 2004. "Communication, reputation, and punishment in sequential bargaining experiments." *Journal of Institutional and Theoretical Economics JITE* 160(4):576-606.

Brown, T.C., I. Ajzen, and D. Hrubes. 2003. "Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation." *Journal of Environmental Economics and Management* 46(2):353-361.

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

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

Carlsson, F., P. Frykblom, and C. Johan Lagerkvist. 2005. "Using cheap talk as a test of validity in choice experiments." *Economics letters* 89(2):147-152.

Carson, R.T. 2011. *Contingent Valuation: A Comprehensive Bibliography and History*. Northampton, MA: Edward Elgar.

---. 2012. "Contingent Valuation: A Practical Alternative When Prices Aren't Available." *The Journal of Economic Perspectives* 26(4):27-42.

Champ, P.A., R. Moore, and R.C. Bishop. 2009. "A comparison of approaches to mitigate hypothetical bias." *Agricultural & Resource Economics Review* 38(2):166.

Charness, G., and B. Grosskopf. 2004. "What makes cheap talk effective? Experimental evidence." *Economics letters* 83(3):383-389.

Cummings, R.G., and L.O. Taylor. 1999. "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method." *The American Economic Review* 89(3):649-665.

Day, B.H., I.J. Bateman, R.T. Carson, D. Dupont, J.J. Louviere, S. Morimoto, R. Scarpa, and P. Wang. 2012. "Ordering effects and choice set awareness in repeat-response stated preference studies." *Journal of Environmental Economics and Management* 63(1):73-91.

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

Doyon, M., L. Saulais, B. Ruffieux, and D. Bweli. 2015. "Hypothetical Bias for Private Goods: Does Cheap Talk Make a Difference?" *Theoretical Economics Letters* 5(06):749.

Gao, Z., L.A. House, and J. Xie. 2015. "Online Survey Data Quality and Its Implication for Willingness-to-Pay: A Cross-Country Comparison." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*.

Gracia, A., J. Barreiro-Hurlé, and B.L. Galán. 2014. "Are local and organic claims complements or substitutes? A consumer preferences study for eggs." *Journal of Agricultural Economics* 65(1):49-67.

Guo, X., H. Liu, X. Mao, J. Jin, D. Chen, and S. Cheng. 2014. "Willingness to pay for renewable electricity: A contingent valuation study in Beijing, China." *Energy Policy* 68:340-347.

Gyrd-Hansen, D. 2016. "The role of the payment vehicle in non-market valuations of a health care service: willingness-to-pay for an ambulance helicopter service." *Health Economics, Policy and Law* 11(01):1-16.

Hausman, J. 2012. "Contingent Valuation: From Dubious to Hopeless." *The Journal of Economic Perspectives* 26(4):43-56.

Hensher, D.A., and W.H. Greene. 2011. "Valuation of travel time savings in WTP and preference space in the presence of taste and scale heterogeneity." *Journal of Transport Economics and Policy*:505-525.

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

Hole, A.R., and J.R. Kolstad. 2012. "Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment." *Empirical Economics* 42(2):445-469.

Jacquemet, N., R.-V. Joule, S. Luchini, and J.F. Shogren. 2013. "Preference elicitation under oath." *Journal of Environmental Economics and Management* 65(1):110-132.

Jones, M.S., L.A. House, and Z. Gao. 2015. "Respondent Screening and Revealed Preference Axioms: Testing Quarantining Methods for Enhanced Data Quality in Web Panel Surveys." *Public Opinion Quarterly*:nfv015.

Kallas, Z., and J.M. Gil. 2015. "Do the Spanish want biodiesel? A case study in the Catalan transport sector." *Renewable Energy* 83:398-406.

Ladenburg, J., O. Bonnichsen, and J.O. Dahlgaard. 2011. "Testing the effect of a short cheap talk script in choice experiments." *Danish J. Econ* 149:25-54.

Lagerkvist, C.J., S. Hess, J. Okello, and N. Karanja. 2013. "Consumer willingness to pay for safer vegetables in urban markets of a developing country: The case of Kale in Nairobi, Kenya." *The Journal of Development Studies* 49(3):365-382.

Lim, K.H., and W. Hu. 2015. "How Local Is Local? A Reflection on Canadian Local Food Labeling Policy from Consumer Preference." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*.

Lindhjem, H., and S. Navrud. 2011. "Are Internet surveys an alternative to face-to-face interviews in contingent valuation?" *Ecological economics* 70(9):1628-1637.

List, J.A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *The American Economic Review* 91(5):1498-1507.

Loomis, J.B. 2014. "Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys." *Journal of Agricultural and Resource Economics* 39(1):34-46.

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

Lusk, J.L., and D. Hudson. 2004. "Effect of monitor-subject cheap talk on ultimatum game offers." *Journal of Economic Behavior & Organization* 54(3):439-443.

Mahieu, P.-A. 2010. "Does gender matter when using cheap talk in contingent valuation studies." *Economics Bulletin* 30(4):2955-2961.

Marta-Pedroso, C., H. Freitas, and T. Domingos. 2007. "Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews." *Ecological economics* 62(3):388-398.

Moser, R., R. Raffaelli, and S. Notaro. 2014. "Testing hypothetical bias with a real choice experiment using respondents' own money." *European Review of Agricultural Economics* 41(1):25-46.

Murphy, J.J., T. Stevens, and D. Weatherhead. 2005. "Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism?" *Environmental and Resource Economics* 30(3):327-343.

Nielsen, J.S. 2011. "Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews." *Resource and energy economics* 33(1):119-129.

Olsen, S.B. 2009. "Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods." *Environmental and Resource Economics* 44(4):591-610.

Ortega, D.L., H.H. Wang, N.J.O. Widmar, and L. Wu. 2014. "Chinese producer behavior: Aquaculture farmers in southern China." *China Economic Review* 28:17-24.

Özdemir, S., F.R. Johnson, and A.B. Hauber. 2009. "Hypothetical bias, cheap talk, and stated willingness to pay for health care." *Journal of health economics* 28(4):894-901.

Penn, J., and W. Hu. 2016. "Understanding Hypothetical Bias: An Enhanced Meta-Analysis." *In review*.

Poe, G.L., J.E. Clark, D. Rondeau, and W.D. Schulze. 2002. "Provision point mechanisms and field validity tests of contingent valuation." *Environmental and Resource Economics* 23(1):105-131.

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.

Scarpa, R., M. Thiene, and K. Train. 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.

Silva, A., R.M. Nayga Jr, B.L. Campbell, and J.L. Park. 2012. "Can perceived task complexity influence cheap talk's effectiveness in reducing hypothetical bias in stated choice studies?" *Applied Economics Letters* 19(17):1711-1714.

---. 2011. "Revisiting Cheap Talk with New Evidence from a Field Experiment." *Journal of Agricultural and Resource Economics* 36(2):280.

Smyth, J.D., D.A. Dillman, L.M. Christian, and M. Mcbride. 2009. "Open-Ended Questions in Web Surveys Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality?" *Public Opinion Quarterly* 73(2):325-337.

Toepoel, V., and M.P. Couper. 2011. "Can verbal instructions counteract visual context effects in web surveys?" *Public Opinion Quarterly* 75(1):1-18.

Toepoel, V., and D.A. Dillman. 2011. "Words, numbers, and visual heuristics in web surveys: Is there a hierarchy of importance?" *Social Science Computer Review* 29(2):193-207.

Tonsor, G.T., and R.S. Shupp. 2011. "Cheap Talk Scripts and Online Choice Experiments:"Looking Beyond the Mean"." *American Journal of Agricultural Economics* 93(4):1015-1031.

Tourangeau, R. 1984. "Cognitive science and survey methods." *Cognitive aspects of survey methodology: Building a bridge between disciplines*:73-100.

Train, K., and M. Weeks. 2005. *Discrete choice models in preference space and willingness-to-pay space*: Springer.

Train, K.E. 2009. *Discrete Choice Methods with Simulation*. 2 ed: Cambridge University Press.

US Census Bureau. 2015. *2014 American Community Survey 1-Year Estimates.*

Van Wezemael, L., V. Caputo, R.M. Nayga, G. Chryssochoidis, and W. Verbeke. 2014. "European consumer preferences for beef with nutrition and health claims: A multi-country investigation using discrete choice experiments." *Food Policy* 44:167-176.

Vista, A.B., R.S. Rosenberger, and A.R. Collins. 2009. "If you provide it, will they read it? Response time effects in a choice experiment." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 57(3):365-377.

Windle, J., and J. Rolfe. 2011. "Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys." *Economic Analysis and Policy* 41(1):83-97.

Yao, R.T., R. Scarpa, J.A. Turner, T.D. Barnard, J.M. Rose, J.H. Palma, and D.R. Harrison. 2014. "Valuing biodiversity enhancement in New Zealand's planted forests: Socioeconomic and spatial determinants of willingness-to-pay." *Ecological economics* 98:90-101.

Zanoli, R., R. Scarpa, F. Napolitano, E. Piasentier, S. Naspetti, and V. Bruschi. 2013. "Organic label as an identifier of environmentally related quality: A consumer choice experiment on beef in Italy." *Renewable Agriculture and Food Systems* 28(01):70-79.

Figure 1:

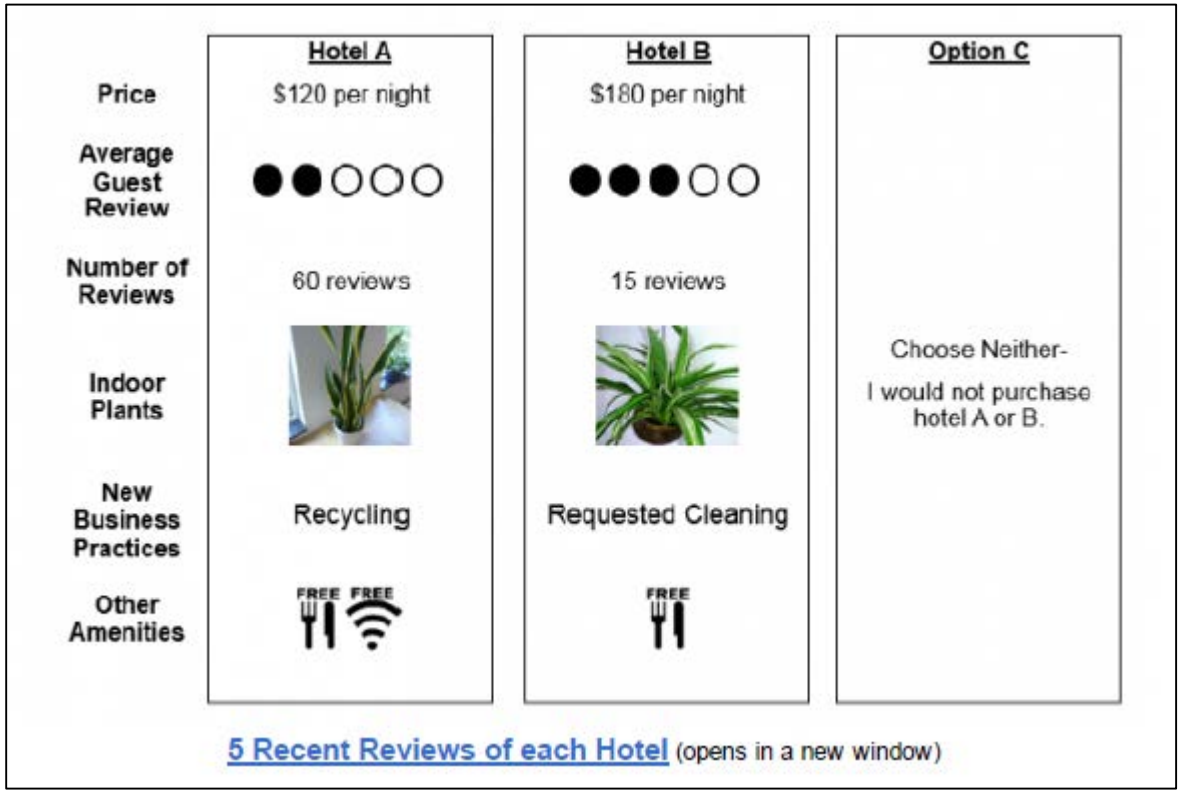
Hotel A
Hotel B
Option C
Price
$120 per night
$180 per night
Average Guest Review
Number of Reviews
60 reviews
15 reviews
Indoor Plants
Choose Neither-
I would not purchase hotel A or B.
New Business Practices
Recycling
Requested Cleaning
Other Amenities
FREE FREE
FREE
5 Recent Reviews of each Hotel (opens in a new window)

Table 1: Choice Experiment Attributes

Attribute (# of levels)	Description
Price (6)	$60, $90, $120, $150, $180, and $210 per evening
Guest Review (4)	Poor: The average online review is a 2 on a 5 point scale
	Average: The average online review is a 3 on a 5 point scale
	Good: The average online review is a 4 on a 5 point scale
	Excellent: The average online review is a 5 on a 5 point scale
Number of Reviews (2)	15 reviews, 60 reviews
Indoor Plant (3)	No live plant in the room
	Spider plan**t**, described as: "Plant A (*chlorphytum comosum*): cleans the air of common indoor pollutants."
	Snake plant, described as: "Plant B (*sansevieria trifasciata*): cleans twice as many common indoor pollutants as Plant A."
Living Wage (2)	Fulltime hotel employees earn higher than the minimum wage to support their families beyond basic subsistence.
Recycling (2)	Recycle bins are placed throughout the entire hotel next to existing trash bins to divert cardboard, plastic, glass and other reusable materials from entering the waste stream.
Room Cleaning upon request only	To save water and energy, room cleaning services will only occur upon guest request.
Breakfast (2)	Free breakfast
Wi-Fi (2)	Free in-room Wi-Fi
Reports of Bed Bugs (3)	0 of 5 recent reviews mention bed bugs
	1 of 5 recent reviews mention beg bugs
	2 of 5 recent reviews mention bed bugs

Table 2: Descriptive Statistics across Treatment Groups

	Overall Sample	Cheap Talk	Attention-Check	Correct Attention-Check	Video	Correct Video Attention-Check	Treatment Differences[2]
n	2068	687	735	364	646	501	
Age[3]	48.6	48.0	48.9	48.7	49.0	48.8	--
Household Income[3] ($thousands)	$84.0	$83.1	$82.8	$88.6	86.5	$90.5	--
% Female	47.8	47.7	48.6	50.5	47.1	47.5	--
% Married	62.6	62.0	61.6	59.1	64.4	64.8	--
% Minor Children	29.9	28.1	28.9	28.5	33.0	32.9	--
% White	80.6	80.6	82.0	82.3	78.9	80.4	--
% Business traveler[6]	37.7	38.0	37.0	39.0	38.1	38.7	--
Education							
% High School	12.1	10.4	12.7	8.0	13.3	9.6	Comp v. Correct Comp
% Some College	31.0	31.6	32.7	30.8	28.5	27.4	--
% Bachelor's	33.4	33.7	31.0	33.0	36.0	38.6	--
% Advanced	22.6	22.8	23.1	28.0	21.7	24.2	--
Employment Status[5]							
% Full Time	45.5	48.0	43.8	44.5	44.8	45.7	--
% Student	4.4	5.2	4.4	5.2	3.4	3.6	--
% Retired	21.5	19.5	23.6	22.0	21.1	20.8	

[2]Statistical tests for differences across treatments. "--" indicates no significant differences for any pairwise comparison. Significant differences are listed for a p-value<.05 based on a t-test of means for Household Income and Age, and a t-test of differences in proportions for the remaining variables.
[3]Based on the midpoint of responses.
[4]Median.
[5]Categories are not mutually exclusive; respondents could list multiple employment statuses.
[6]Defined as at least 7 nights in hotels for business/work-related purposed in the past year.

Table 3: WTP Space Mixed logit Model Results

	Standard CT		Attention CT		Correct Attention CT		Video CT		Correct Video CT	
	WTP Estimate (Std. Err)	WTP Std. Dev. (Std. Err)	WTP Estimate (Std. Err)	WTP Std. Dev. (Std. Err)	WTP Estimate (Std. Err)	WTP Std. Dev. (Std. Err)	WTP Estimate (Std. Err)	WTP Std. Dev. (Std. Err)	WTP Estimate (Std. Err)	WTP Std. Dev. (Std. Err)
Opt-Out	-80.48*** (6.85)	74.83*** (4.68)	-85.57*** (6.24)	72.27*** (4.64)			-81.29*** (5.98)	66.21*** (4.58)	-79.89*** (6.65)	59.01*** (4.69)
Average	23.63*** (4.48)	25.75*** (8.11)	26.09*** (4.05)	22.92*** (6.07)			24.23*** (3.56)	0.83 (10.3)	23.53*** (4.23)	9.33 (6.47)
Good	69.26*** (4.79)	25.34*** (8.08)	60.43*** (4.23)	-13.47 (9.21)			57.08*** (3.98)	19.93*** (6.82)	58.78*** (4.52)	23.17*** (7.42)
Excellent	75.88*** (4.71)	25.92*** (8.04)	69.25*** (4.19)	19.09** (7.44)			59.86*** (3.92)	11.59* (6.17)	63.28*** (4.6)	14.28** (6.67)
Number of Reviews	-0.19*** (0.06)	0.00 (0.19)	-0.14** (0.06)	-0.37*** (0.12)			-0.04 (0.06)	0.34** (0.14)	-0.06 (0.07)	0.43*** (0.1)
Spider Plant	6.33* (3.70)	9.94 (8.04)	5.81* (3.42)	-14.79** (5.96)			11.28*** (3.13)	10.6** (4.68)	11.18*** (3.61)	5.02 (8.53)
Snake Plant	5.60 (3.56)	1.71 (5.09)	6.48* (3.38)	-12.41** (5.82)			8.26*** (3.1)	8.49* (4.47)	8.38** (3.56)	5.45 (7.23)
Living Wage	13.71*** (2.7)	3.62 (9.78)	9.56*** (2.58)	22.55*** (4.44)			7.63*** (2.25)	17.78*** (3.55)	7.44*** (2.56)	2.27 (7.11)
Recycling	16.47*** (2.61)	0.44 (4.49)	17.57*** (2.41)	-2.46 (4.31)			8.53*** (2.25)	3.95 (3.64)	10.12*** (2.59)	1.59 (4.73)
Room Cleaning by Request	6.85*** (2.57)	7.22 (6.19)	10.73*** (2.38)	9.03* (5.37)			6.09*** (2.23)	6.32 (7.37)	3.80 (2.83)	16.57*** (5.41)
Complimentary Breakfast	11.15*** (2.94)	14.88*** (4.92)	10.07*** (2.73)	15.4*** (4.07)			5.01** (2.43)	5.28 (6.33)	5.26* (2.82)	5.33 (3.82)
Complimentary In-Room Wi-Fi	32.02*** (2.78)	15.32*** (4.94)	23.63*** (2.54)	-22.88*** (4.2)			27.27*** (2.45)	16.82*** (4.88)	27.78*** (2.81)	17.68*** (3.84)
Reports of Bed Bugs	21.80*** (2.43)	40.65*** (2.97)	19.81*** (2.13)	32.38*** (2.69)			17.58*** (2.02)	25.25*** (2.92)	18.55*** (2.29)	27.8*** (2.89)
Log Likelihood	4786.433		5144.884		2563.433		4348.647		3338.82	
N choice sets	5496		5880		2912		5168		4008	
Note: *, **, and * indicate statistical significance at a p-value less than .01, .05, and .1, respectively.**										

Table 3: Pooled WTP Space Mixed Logit Model Results

Standard CT-Video CT			Standard CT-Attention-Check CT			Combinatorial Test
	WTP Estimate (Std. Err.)	p-value		WTP Estimate (Std. Err.)	p-value	
Video*Opt-Out	-22.90 (11.71)	0.051	**Attention*Opt-Out**	-16.38 (10.9)	0.133	
Video*Average	0.96 (6.26)	0.878	**Attention*Average**	-1.01 (6.03)	0.468	
Video*Good	-1.45 (6.51)	0.824	**Attention*Good**	-12.23 (6.1)	0.868	A,C, D
Video*Excellent	-5.29 (6.78)	0.436	**Attention*Excellent**	-8.32 (6.15)	0.045	A, C, D, E, F
Video*Number of Reviews	0.1 (0.1)	0.304	**Attention*Number of Reviews**	0.01 (0.09)	0.176	
Video*Spider Plant	8.59 (5.54)	0.121	**Attention*Spider Plant**	-2 (5.13)	0.943	
Video*Snake Plant	4.9 (5.38)	0.362	**Attention*Snake Plant**	0.59 (5.03)	0.697	
Video*Living Wage	-2.55 (3.96)	0.52	**Attention*Living Wage**	-1.79 (3.85)	0.907	D
Video*Recycling	-5.77 (3.86)	0.135	**Attention*Recycling**	-0.15 (3.67)	0.642	C, D, E, F
Video*Room Cleaning by Request	-1.03 (3.77)	0.785	**Attention*Room Cleaning by Request**	3.63 (3.48)	0.968	E
Video*Free Breakfast	-3.27 (4.3)	0.447	**Attention*Free Breakfast**	-1.09 (3.97)	0.298	C
Video*Free In-Room Wi-Fi	2.93 (3.91)	0.453	**Attention*Free In-Room Wi-Fi**	-7.00 (3.67)	0.784	A, B, F
Video*Reports of Bed Bugs	-0.86 (3.37)	0.798	**Attention*Reports of Bed Bugs**	-0.62 (3.11)	0.056	
Log Likelihood	9046.486					
N choice sets	10664					

Note: Each of the following is based on the Combinatorial test by Poe et al. (2005) for p-value <.1
A: indicates a significant difference between the Standard CT and Attention-Check treatments
B: indicates a significant difference between the Standard CT and Correct Attention-Check treatments
C: indicates a significant difference between the Standard CT and Video treatments
D: indicates a significant difference between the Standard CT and Correct Video treatments
E: indicates a significant difference between the Attention-Check and Video treatments
F: indicates a significant difference between the Correct Attention-Check and Correct Video treatments