Hypothetical Bias Mitigation in Choice Experiments:

Cheap Talk Fades with Repeated Choices while Honesty Priming is Ineffective Online

Gregory Howard
Department of Economics, East Carolina University

Brian Roe
Department of AED Economics, Ohio State University

Erik Nisbet
School of Communication, Ohio State University

Abstract: We design a choice experiment comparing policies that reduce agricultural nutrient pollution and harmful algal blooms in Lake Erie and administer it to Ohio residents using an online survey panel. We compare two treatments that have been found to mitigate hypothetical bias, cheap talk and honesty priming. We find greater sensitivity to price among respondents during choices made immediately following the cheap talk intervention. As additional choices are made, price sensitivity diminishes and eventually matches that of respondents in the control treatment. We find this effect in both our online choice experiments and among respondents to face-to-face choice experiments conducted by de-Magistris, Gracia and Nayga (2013, DNG). Our online implementation of an honesty priming intervention yields no significant change in price sensitivity compared to a control. While DGN (2013) implement an honesty priming intervention that fully mitigates hypothetical bias in a face-to-face setting, we show this effect is also transient, and in later choice exercises we cannot reject the null hypothesis of equality between honesty priming and the control. Our results suggest additional work is required to adapt priming interventions for online settings and to extend the effectiveness of popular hypothetical bias mitigation techniques to later rounds in choice experiments.

There is evidence that responses to hypothetical questions and situations may deviate systematically from responses observed in real world interactions. As such, data collected in low- or no-incentive environments may fail to achieve the goal of accurately representing real-world behavior. This deviation, hypothetical bias, affects a wide range of behaviors and preferences. For example, research has found that individuals tend to exhibit greater risk tolerance (Holt and Laury 2005; Howard 2014) and greater valuations for a wide range of both private and public goods (Harrison and Rutstrom 2008; List and Gallet 2001) when posed hypothetical rather than real incentives. This finding poses a particular challenge to the valuation of environmental amenities. In many cases, it is difficult or impossible to adequately incentivize responses to such valuation questions. For instance, studies that use a referendum format to estimate willingness to pay (WTP) can seldom implement policy based on the outcome of the study referendum.

The difficulty of devising incentive-compatible valuation exercises, coupled with the attractiveness of using hypothetical questions from a cost and convenience standpoint, has led researchers to examine the possibility of mitigating or eliminating hypothetical bias when hypothetical techniques are used. This literature has developed and tested several possible remedies, with varying levels of effectiveness. Further, effectiveness may depend on the mode of preference elicitation. Hypothetical questions can be posed via lab experiment, field experiment, phone survey, mail survey, and internet survey. Each mode exerts different levels of researcher control over the respondent's focus and time commitment. Because of this, it is unclear whether an intervention that eliminates hypothetical bias in one mode will have a similar effect in others. In particular internet surveys, which have grown in popularity in the past decade, may command markedly different levels of respondent time and attention than controlled lab or field experiments.

This paper makes two unique contributions to the literature on hypothetical bias in elicitation. First, we compare the effects of two hypothetical bias treatments in the context of an online survey methodology. One treatment, a “cheap talk” script (Cummings and Taylor 1999), has been shown effective in multiple research contexts, including online surveys (Tonsor and Shupp 2011). The other treatment, honesty priming, has been shown effective in lab experiments (de-Magistris, Garcia and Nayga 2013) but has not yet been tested in an online context.

It is common practice to present each respondent with multiple choice exercises, often as many as sixteen (Louviere, Hensher and Swait 2000). While some research has examined how presenting a greater number of choice sets may influence WTP estimates (Bech, Kjaer and Lauridsen 2011), to our knowledge no study has estimated the effectiveness of hypothetical bias mitigation tools as more choice exercises are added. Our second contribution is to analyze whether the impact of our interventions, cheap talk and honesty priming, is persistent or transient. As most researchers present each respondent with multiple choice exercises, it is possible that the effectiveness of an intervention may be strongest for early exercises and dissipate as time passes and later choices are made. This is a crucial point of interest that is addressed for the first time in this study. Furthermore, because cheap talk is an overt attempt to change behavior while honesty priming is better defined as inconspicuous, ours is the first study that compares whether overt and inconspicuous treatments differ in persistence.

We find the cheap talk intervention reduces WTP for improved environmental outcomes by about 50% relative to the control, which is in line with estimates of hypothetical bias in the literature. This effect is entirely driven by increases in the marginal disutility of price; there are no differences in the marginal utility of environmental improvements between the treatments and control. We find no significant difference between the honesty priming treatment and the control. This result is in contrast to the finding of de-Magistris, Garcia and Nayga (2013, hereafter DGN), in which honesty priming fully eliminates hypothetical bias in a face-to-face lab setting.

While we find an aggregate effect of the cheap talk treatment, our exercise-level analysis reveals the effectiveness of cheap talk is transient; cheap talk increases price sensitivity relative to the control in the initial choice exercises but not in later choice exercises. This finding is corroborated in a similar exercise-level analysis using data from DGN's in-person lab experiments for both cheap talk and honesty priming treatments.

The rest of the paper proceeds as follows. Section 1 reviews the relevant literature. Section 2 describes the survey and data. Section 3 outlines the empirical model, Section 4 presents results and Section 5 provides discussion and concludes.

1. Literature

Numerous methods have been developed to eliminate hypothetical bias in choice exercises.[1] The majority of these methods fall under the category of "ex-ante" fixes, meaning they address the problem during the elicitation process.[2] One method increases the

[1] For a description of the competing theories explaining the source of hypothetical bias, see Loomis (2011, 2014).

[2] Another vein of the literature addresses the problem using "ex-post" fixes in which researchers elicit responses that may suffer from hypothetical bias and adjust the analysis to account for this potential bias in one of several ways. Most ex-post fixes use a certainty question following the choice exercise, though the specific use of this question in subsequent analysis varies (Blumenschein et al. 2008; Champ, Moore and Bishop 2009; Moore *et al*. 2010). Still another ex-post fix draws not on certainty questions, but instead involves calibrating hypothetical responses using bias-correction factors derived from the literature (Fox *et al*. 1998).

consequentiality of the questions posed to respondents by convincing them that the questions they face are not purely hypothetical, but may have real consequences (Carson and Groves 2007). This is often achieved by informing respondents that the results of the survey will be used by policy makers to determine future policy. This tactic may curb hypothetical bias, but can also induce strategic responses that do not reflect underlying preferences (Satterhwaite 1975) and may require deception when researchers cannot legitimately claim that policy makers will use the survey to inform their decisions.

Another ex-ante approach is the presentation of a cheap talk script prior to the choice exercise. While cheap talk scripts vary in length and specifics, they generally implore the respondent to treat the following hypothetical choice as if it were a real choice involving real money. Longer scripts also educate the respondent on the tendency of survey-takers to overestimate WTP and suggest reasons for this tendency. Several studies have found that cheap talk scripts eliminate hypothetical bias (Cummings and Taylor 1999; Silva *et al.* 2011), mitigate hypothetical bias (Moser, Raffaelli and Notaro 2014; Champ, Moore and Bishop 2009; de-Magistris and Pascucci 2014), or induce differential effects based on respondent attributes and context (List 2001; Silva *et al.* 2012). Other studies document a decrease in WTP but cannot assess the degree of bias mitigation as no fully incentivized treatment is conducted (Whitehead and Cherry 2007; Tonsor and Shupp 2011; Carlsson, Frykblom and Lagerkvist 2005). Blumenschein *et al.* (2008) find no effect from cheap talk, while Morrison and Brown (2009) find that cheap talk overcorrects for hypothetical bias. While the full story is nuanced, it suffices to say that findings for cheap talk are generally positive but mixed. More recently, researchers have utilized an "oath script," which asks respondents to swear an oath to answer truthfully before being presented with the choice exercise. This method has also been effective at mitigating hypothetical bias (Jacquemet *et al.* 2013; Carlsson *et al.* 2013; de-Magistris and Pascucci 2014).

Adapting techniques from social psychology, DGN utilize an "honesty priming" exercise. While the cheap talk and oath scripts overtly encourage respondents to answer honestly, honesty priming engages the respondent in a simple task that endeavors to subconsciously prime subjects for honesty. Several tasks have been demonstrated effective at priming for truthfulness, including the scrambled-sentence method used by DGN (Chartrand and Bargh 1996). DGN show that WTP in the honesty priming treatment is significantly less than in the hypothetical baseline treatment and that the honesty priming treatment is not significantly different from the incentivized treatment.

In addition to this work on hypothetical bias, our research also draws on a literature that examines how different survey modes (face-to-face, mail, phone, internet, etc.) may influence respondent behavior. Lindhjem and Navrud (2011) provide both a theoretical examination of survey mode effects and a thorough review of the literature comparing online surveys with other survey modes. The majority of studies find demographic differences between internet survey respondents and respondents from other modes, though the use of internet panels and data weighting techniques can mitigate this issue. The literature comparing WTP between internet and other survey modes is mixed, with many studies failing to reject equality of estimates (Banzhaf *et al*. 2006; Nielsen 2011; Olsen 2009; Windle and Rolfe 2011) and many others rejecting equality (Canavari, Nocella and Scarpa 2005; Marta-Pedroso, Freitas and Domingos 2007; Bell, Huber and Viscusi 2011). To our knowledge, only one study has examined the effectiveness of cheap talk scripts in an internet survey setting (Tonsor and Shupp 2011) and no studies have performed a similar examination of honesty priming.

Several studies have considered how early and late choice exercises may differ from each other. Carlsson *et al.* (2012) present 16 choice exercises to subjects, with the last eight being identical to the first eight. They find smaller error variances for the second group of exercises. Further, they find most of the differences that exist between the first and second set of eight choices come from changes in response to the first choice exercise. Ladenburg and Olsen (2008) identify starting point bias in a choice experiment and note that the effect is transient, meaning it fades in later choice exercises. Bateman *et al.* (2008) find that the effects of anchoring dissipate in later choice exercises, as do inconsistencies between single bound and double choice formats. Brouwer *et al.* (2010) administer a choice experiment with certainty follow-up questions after each exercise. They test whether reported certainty changes from early to late exercises and find no evidence of certainty changes when they control for demographics and choice attributes. While each of these studies has explored how respondent behavior may fluctuate over multiple choice exercises, and some studies have affirmed that respondent preferences are not invariant to choice experiment length, no study to this point has considered whether this phenomenon exists in the realm of hypothetical bias mitigation.

2. Survey and Data

We use an online survey with a between-subjects approach to compare two treatment tasks (cheap talk and honesty priming) with a neutral priming control task. In both treatments and the control, these tasks immediately preceded a sequence of five choice exercises. In our priming task, which follows Vinski and Watter (2012), respondents were given six word sets. Each word set contained one target word and three additional words. One of the additional words was a synonym of the target word, while the others additional words had similar meanings without being synonyms. Respondents were asked to enter the synonym in the text space provided. In the honesty priming task, four of the six target words (factual, honest, candid, and sincere) were intended to prime for honesty. In the neutral priming task, all six target words did not prime for honesty, but instead were chosen to avoid priming for any specific attitude or behavior. This honesty priming task differed from the scrambled sentence task used by DGN, although both tasks have been shown effective in priming honesty and truthful revelation of information (Chartrand and Bargh 1996; Rasinski *et al*. 2004). The cheap talk treatment presented the respondent with a 322 word script[3] that resembles other scripts used in the literature. In order to maintain a similar level of interaction between the control and both treatments, respondents are asked to provide a brief summary of the script in the cheap talk treatment.

The choice experiment asks respondents to rank three options: a status-quo option and two programs aimed at reducing harmful algal blooms (HABs) in Lake Erie by reducing nutrient

[3] The full cheap talk script presented to respondents: "Later in this survey, you will be presented with a hypothetical choice involving money. No one will actually be paid money based on the decision you make, but you are asked to make the decision as though it would result in the actual payment.
"Studies show that people tend to act differently when they face hypothetical decisions. In other words, they say one thing and do something different. We call this a 'hypothetical bias.' For example, in a recent study, several different groups of people made decisions just like the one you are about to make. Payment was real for one group and hypothetical for the other group, as it will be for you. The results of these studies were that on average, more people expressed a willingness to pay money in the hypothetical group than in the real group.
"How can we get people to think about their decision in a hypothetical situation like they think in a real situation? I think that when we hear about a situation that involves doing something that is basically good, for example helping people in need, improving environmental quality, or anything else, our basic reaction in a hypothetical situation is to think: sure, I would do this. I really would spend the money; I really, really, think I would.
"But when the situation is real, and we would actually have to spend our money, we think a different way. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money, we think about our options: If I spend money on this, that's money I cannot spend on other things. So, when the payment is real, we act in a way that takes into account the limited amount of money we have. We make the decision while realizing that we just don't have enough money to do everything we might like to do."

pollution from agriculture. Changes in Lake Erie's watershed during the past three decades reveal tight linkages between lake health and upstream behavior that influences nutrient dynamics. During the 1970s through mid-1990s, programs targeting point source phosphorus abatement resulted in steady reductions in HABs and improved water clarity (Makarewicz and Bertram 1991; Ludsin et al. 2001). However, since the mid-1990s, Lake Erie has entered a transitional state, due to simultaneous human and environmental influences (Matisoff and Ciborowski 2005). Recently, phosphorus from agricultural runoff has been identified as the dominant source of resurgent HABs (Ohio EPA 2010) and, in the summer of 2014, after the implementation of this survey, HABs caused the municipal water supply for the city of Toledo to become unsafe, requiring more than 400,000 residents to find other sources of water for drinking and bathing for more than two days. Hence, knowledge of Ohio residents' willingness to pay for policies that alter farm nutrient runoff and subsequent HAB-induced losses of ecosystem services is in great demand by policymakers.

Each program has three attributes that are varied: program cost, program effectiveness, and program details. Program cost is the expected annual cost to the respondent's household and took one of four levels ($19, $34, $71, and $102) for each new proposed program. Program effectiveness was captured using three different attributes (fish kills, annual beach closure and water quality advisory days, and a satellite image of the expected annual HAB). New programs took one of three possible levels, which are measured as the percentage reduction of these undesirable outcomes from the baseline status quo outcome (levels are 10%, 20% and 50% reduction).[4] Program details explain the manner in which funds will be collected and the way in which the program will reduce nutrient pollution. This attribute takes one of five possible levels

[4] In our design, all three indicators of program effectiveness move in unison. This means that a general value of program effectiveness can be identified, but not the value of individual program effectiveness indicators.

(Payment for ecosystem service (PES) programs funded by income tax, PES programs funded by sales tax, regulation funded by income tax, regulation funded by sales tax, and a fertilizer tax). With the fertilizer tax, respondents were informed that the annual cost to the household is based on higher food prices.

The status quo was labeled the "current program." This program was given the same program details (PES funded by sales tax) and effectiveness (90 annual beach closure and water quality advisory days, 5,467 fish die from fish kills annually) for all choice exercises. Additionally subjects were randomly assigned one of two possible program costs for the status quo ($2 or $5). While there was between-subject variation in current program cost, there was no within-subject variation. This variation of the status quo cost was necessary to allow for evaluation of cost as a continuous variable while still controlling for alternative specific constants. Figure 1 displays an example choice exercise.

Experimental design was determined using several experimental design macros (%mktruns, %mktex, %mktroll, %choiceff, and %mktblock) available in SAS 9.3. Before beginning the design, we developed and applied a restriction macro similar to those found in Kuhfeld (2010) in order to eliminate choice sets with a dominated program.[5] The resulting design of 20 choice exercises achieves a relative D-efficiency of 90%. The design of 20 was converted into four blocks of five choice exercises each using an efficient choice blocking macro. The order of choice exercise presentation was randomized, ensuring that the order of presentation is not confounded with specific attribute levels in our analysis.

[5] Dominated programs were based on program cost and effectiveness, since we made no ex-ante assumptions about the relative desirability of different program detail options.

Figure 1: Choice Exercise Example

	Program A	Program B	Current Program
Annual Cost to your Household	$102	$71	$5
Fish Kills	2,733 fish die annually	4,921 fish die annually	5,467 fish die annually
Number of Annual Beach Closure and Water Quality Advisory Days	45 annual beach closure and water quality advisory days	81 annual beach closure and water quality advisory days	90 annual beach closure and water quality advisory days
Annual Expected Size of Lake Erie Algae Bloom (Satellite Photo)			
Program Details	The government will offer Payment for Ecosystem Service (PES) programs. Farmers who choose to enroll in PES programs will be compensated for implementing practices on their farms that reduce nutrient runoff. The program is funded using **state income taxes**. Given your stated income level, the annual cost of this program to your household will be $102	The government will introduce a new tax on fertilizers. This will reduce nutrient use and nutrient runoff in the watershed. The tax will indirectly affect households through higher food prices. Given your stated income level, the annual cost of this program to your household will be $71	The government will offer Payment for Ecosystem Service (PES) programs. Farmers who choose to enroll in PES programs will be compensated for implementing practices on their farms that reduce nutrient runoff. The program is funded using **state sales taxes**. Given your stated income level, the annual cost of this program to your household will be $5

Please provide a ranking for the above programs where 1=Best, 2=Middle, and 3=Worst

	1	2	3
Program A	○	○	○
Program B	○	○	○
Current Program	○	○	○

The data were collected from a survey of Ohio residents in March 2014. Respondents were recruited from six different online panels using Qualtrics. After removing respondents who failed to complete two "focus tests" embedded in the survey,[6] our sample consisted of 1,209 responses which were representative of the general population of Ohio in several demographic indicators (gender, age and proportion of the population black vs. nonblack), although the sample is skewed toward individuals with more formal education (35.8% of the sample has completed a four-year college degree, compared with 28% in Ohio).

3. Empirical Model

We utilize a random utility model. Indirect utility for individual *i* associated with program *j* is given by the following equation:

$$U_{ij} = V_{ij} + e_{ij}, \tag{1}$$

where U_{ij} is latent or unobserved utility, V_{ij} is observable utility, and e_{ij} is the random or unobservable portion utility for each choice. We further specify the utility of a program is a function of a vector of program attributes X_{ij}:

$$U_{ij} = \beta X_{ij} + e_{ij}. \tag{2}$$

Assuming that errors are i.i.d. and follow a type 1 extreme value distribution, the probability that individual *i* will select program *j* as the best is given by

[6] The purpose of these questions is to identify respondents who are not carefully reading and completing the survey. An example of this type of question is as follows: "Sickle cell anemia is simply a different name for malaria. We are checking to see how closely people follow directions. Please select "Not Sure" for this question." Any respondent who does not select "Not Sure" fails the focus test and is removed from the dataset.

$$Pr_{\mathrm{ij}} = \frac{exp(\beta X_{ij})}{\sum_{j=1}^{J} exp(\beta X_{ij})}. \tag{3}$$

This is the standard conditional logit model. Variations of this standard model that account for individual-level heterogeneity, including mixed logit models, latent class models, and attribute non-attendance models, have become increasingly popular in recent years. While these models are useful, and indeed vital, in addressing certain research questions, we contend that our focus on 1) the aggregate impacts of cheap talk and honesty priming and 2) whether and how these aggregate impacts change over the course of a series of made choices does not substantially benefit from the use of heterogeneous preference models.

Table 1 gives a full description of the variables used in our analysis. In our baseline estimation we include only program attributes in the independent variable matrix X_{ij},

$$U_{ij} = \beta_1 Price_{ij} + \beta_2 Effect_{ij} + \beta_3 ProgA_{ij} + \beta_4 ProgB_{ij} + \gamma D_{ij} + e_{ij}, \tag{4}$$

where $Price_{ij}$ and $Effect_{ij}$ capture program price and effectiveness (where positive values indicate percentage reductions in undesirable outcomes), $ProgA_{ij}$ and $ProgB_{ij}$ are alternative-specific constants, and D_{ij} is a series of indicator variables identifying different program detail alternatives.[7] To test the effectiveness of cheap talk (CT) and honesty priming (HP) in influencing willingness to pay, we additionally include the interactions of treatment dummy variables and program[8] attributes:

[7] Our model assumes changes in program effectiveness have a linear impact on utility. This is not self-evident, so as a robustness check we estimated equation (4) and included a squared term for program effectiveness. The coefficient for squared effectiveness is not significant (p value = 0.133).

[8] We do not interact treatments with alternative-specific constants. Bosworth and Taylor (2012) use such an interaction and find that cheap talk can decrease program participation on the extensive margin as well as on the intensive margin. As a robustness check, we interact treatments with alternative-specific constants and find no significant effect at the 95% level (p values 0.229, 0.852, 0.065 and 0.648).

Table 1: Variable Definitions

Variable	Description
Price	Cost of the program to the respondent's household, in dollars
Effect	Effectiveness of the program, in percentage reduction of negative outcomes
Reg*SalesTax	Dummy equal to 1 if the program uses increased sales taxes to fund increased regulation of farmers
Reg*IncomeTax	Dummy equal to 1 if the program uses increased income taxes to fund increased regulation of farmers
PES*SalesTax	Dummy equal to 1 if the program uses increased sales taxes to fund more voluntary farmer PES programs
PES*IncomeTax	Dummy equal to 1 if the program uses increased income taxes to fund more voluntary farmer PES programs
CT	Dummy equal to 1 if respondent received the cheap talk treatment
HP	Dummy equal to 1 if respondent received the honesty priming treatment
Ex*N*	Dummy equal to 1 if the choice exercise was the Nth one faced by the respondent

$$U_{ij} = \beta_1 Price_{ij} + \beta_2 Effect_{ij} + \beta_3 ProgA_{ij} + \beta_4 ProgB_{ij} + \gamma D_{ij} + \alpha_1 CT_i{*}Price_{ij} + \alpha_2 CT_i{*}Effect_{ij} + \rho CT_i{*}D_{ij} + \eta_1 HP_i{*}Price_{ij} + \eta_2 HP_i{*}Effect_{ij} + \theta HP_i{*}D_{ij} + e_{ij}. \qquad (5)$$

Finally, to test whether any impact of our treatments on price sensitivity is persistent or transient, we estimate the following equation:

$$U_{ij} = \beta_2 Effect_{ij} + \beta_3 ProgA_{ij} + \beta_4 ProgB_{ij} + \gamma D_{ij} + \sum_{k=1}^{5} \alpha_k CT_i{*}Price_{ij}{*}Ex_k + \sum_{k=1}^{5} \gamma_k HP_i{*}Price_{ij}{*}Ex_k + \sum_{k=1}^{5} \theta_k Price_{ij}{*}Ex_k + e_{ij}. \qquad (6)$$

This formulation includes ten three-way interactions between price, treatment and exercise and five interactions between price and exercise (which capture the by-exercise price effects for the control group). In all models, the omitted group is the neutral prime control group and the omitted program details category is a fertilizer tax. All models incorporate robust standard errors that are clustered by respondent.

4. Results

Column I of Tables 2 and 3 display the results of our baseline model. This model, detailed in equation (4), estimates a conditional logit using only program attributes as explanatory variables. Columns II, III and IV add to the baseline model by examining the effect of our two treatments on the price, effect, and details attributes as shown in equation (5). Coefficient values are presented in Table 2 and marginal effects are presented in Table 3. As expected, decreasing program cost and increasing program effectiveness each increase the probability of program selection in all models. Additionally, we find that there is a general preference for PES programs funded through increasing sales tax and regulation funded through increases in either sales or income taxes relative to the omitted policy option of a fertilizer tax. There is also a clear preference for any program relative to the "current situation" status quo, even after controlling for program attribute levels.

Our main finding, which is consistent through all models, is the effectiveness of CT and ineffectiveness of HP at increasing price sensitivity relative to the neutral priming control. The impact of price in the CT treatment is larger than for both neutral priming (via *t* test) and HP (via Wald test) at either the 99% or 95% confidence level in each model, while we uniformly cannot reject the null of equal price effects between HP and neutral priming. Specifically, a $10 increase in program price decreases the probability of choosing the program by about 1% (1.0-1.4% across models) for respondents in the control and HP treatment, while a similar price increase has about double the effect (2.2% in all models) in the CT treatment. We find very few treatment effects when examining non-price attributes. Neither treatment has any statistically significant differential effects on program effectiveness. Regarding program details, respondents receiving

Table 2: Estimation Results: Coefficients

Variable	I	II	III	IV
Price	-0.0063** (<0.005)	-0.0045** (<0.005)	-0.0050** (<0.005)	-0.0046** (<0.005)
Effect	0.0259** (<0.005)	0.0259** (<0.005)	0.0269** (<0.005)	0.0266** (<0.005)
Reg*SalesTax	0.1439* (0.027)	0.1443* (0.017)	0.1446* (0.028)	-0.0353 (0.766)
Reg*IncomeTax	0.2104** (<0.005)	0.2124** (<0.005)	0.2123** (0.005)	0.1689 (0.155)
PES*SalesTax	0.2896** (<0.005)	0.2913** (<0.005)	0.2913** (<0.005)	0.2433** (0.009)
PES*IncomeTax	0.1104 (0.163)	0.1048 (0.188)	0.1050 (0.187)	-0.0309 (0.812)
Program A	0.1922* (0.019)	0.1942* (0.017)	0.1945* (0.017)	0.1959* (0.016)
Program B	0.2419** (<0.005)	0.2424** (<0.005)	0.2425** (<0.005)	0.2441** (<0.005)
Price*CT	-	-0.0051** (<0.005)	-0.0045* (0.013)	-0.0050** (0.010)
Price*HP	-	-0.0003 (0.876)	0.0005 (0.771)	-0.0002 (0.908)
Effect*CT	-	-	-0.0015 (0.626)	-0.0001 (0.974)
Effect*HP	-	-	-0.0017 (0.569)	-0.0022 (0.463)
Reg*SalesTax*CT	-	-	-	0.1877 (0.251)
Reg*SalesTax*HP	-	-	-	0.3438* (0.032)
Reg*IncomeTax*CT	-	-	-	-0.0149 (0.930)
Reg*IncomeTax*HP	-	-	-	0.1388 (0.403)
PES*SalesTax*CT	-	-	-	0.1144 (0.379)
PES*SalesTax*HP	-	-	-	0.0386 (0.768)
PES*IncomeTax*CT	-	-	-	0.1248 (0.498)
PES*IncomeTax*HP	-	-	-	0.2809 (0.115)
Wald Test (Price*CT = Price*HP)	-	7.93** (<0.005)	7.85** (0.005)	6.15* (0.013)
Log Likelihood	-6234.47	-6218.23	-6218.03	-6211.39

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 18100 observations come from 1209 respondents, with robust standard errors clustered by respondent. *P*-values reported in parentheses.

Table 3: Marginal effects

Variable	I	II	III	IV
Price	-0.0014** (<0.005)	-0.0010** (<0.005)	-0.0012** (<0.005)	-0.0011** (<0.005)
Effect	0.0060** (<0.005)	0.0060** (<0.005)	0.0062** (<0.005)	0.0061** (<0.005)
Reg*SalesTax	0.0327* (0.022)	0.0328* (0.022)	0.0328* (0.022)	-0.0082 (0.760)
Reg*IncomeTax	0.0475** (<0.005)	0.0479** (<0.005)	0.0479** (<0.005)	0.0383 (0.145)
PES*SalesTax	0.0668** (<0.005)	0.0671** (<0.005)	0.0671** (<0.005)	0.0561** (0.008)
PES*IncomeTax	0.0252 (0.153)	0.0239 (0.177)	0.0239 (0.177)	-0.0071 (0.817)
Program A	0.0440* (0.015)	0.0444* (0.014)	0.0444* (0.014)	0.0447* (0.013)
Program B	0.0552** (<0.005)	0.0552** (<0.005)	0.0552** (<0.005)	0.0556** (<0.005)
Price*CT	-	-0.0012** (<0.005)	-0.0010* (0.013)	-0.0011** (0.010)
Price*HP	-	-0.0001 (0.876)	0.0001 (0.771)	-0.0001 (0.908)
Effect*CT	-	-	-0.0003 (0.626)	-0.0000 (974)
Effect*HP	-	-	-0.0004 (0.569)	-0.0005 (0.463)
Reg*SalesTax*CT	-	-	-	0.0422 (0.238)
Reg*SalesTax*HP	-	-	-	0.0754* (0.023)
Reg*IncomeTax*CT	-	-	-	-0.0034 (0.930)
Reg*IncomeTax*HP	-	-	-	0.0314 (0.393)
PES*SalesTax*CT	-	-	-	0.0261 (0.374)
PES*SalesTax*HP	-	-	-	0.0089 (0.767)
PES*IncomeTax*CT	-	-	-	0.0286 (0.490)
PES*IncomeTax*HP	-	-	-	0.0622 (0.100)

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 18100 observations come from 1209 respondents, with robust standard errors clustered by respondent. *P*-values reported in parentheses.

the HP treatment were more favorable to regulation funded by sales taxes relative to the control, but this is the only significant differential effect of either hypothetical bias treatment on the program details attribute.

Table 4 presents WTP estimates for the control and two treatment groups. These estimates are derived using the model from Column IV of Table 2, in which coefficients for all three program attributes (price, effectiveness, and program details) are allowed to vary by treatment. WTP for an attribute is calculated using the ratio of coefficients between the attribute in question and the price attribute, so that WTP for a one-unit increase in attribute $X = \beta_x/\beta_{price}$. the CT group are half as large as those in HP and the control. This result is driven entirely by the increased sensitivity to price exhibited by CT respondents.

Table 4: Willingness to Pay Estimates from Table 2, Column IV

	Control	**HP**	**CT**
Effect	**$5.81** [3.71, 14.20]	**$5.09** [3.33, 11.02]	**$2.78** [2.17, 3.82]
Reg*SalesTax	-$7.71 [-79.42, 48.61]	**$64.34** [19.57, 158.36]	$15.99 [-8.35, 41.68]
Reg*IncomeTax	$36.88 [-12.90, 160.05]	**$64.16** [14.17, 194.27]	$16.16 [-8.51, 47.80]
PES*SalesTax	**$53.16** [11.54, 165.17]	**$58.81** [19.88, 153.78]	**$37.55** [18.24, 65.44]
PES*IncomeTax	-$6.74 [-82.88, 63.79]	**$52.14** [0.25, 148.40]	$9.86 [-18.68, 39.18]
Tests for Differences in Estimated WTP Distributions			
	Control vs. HP	HP vs. CT	Control vs. CT
Effect	0.3092	0.1372	0.1301
Reg*SalesTax	**0.0286**	**0.0376**	0.1909
Reg*IncomeTax	0.2950	0.0629	0.2339
PES*SalesTax	0.3567	0.1950	0.2506
PES*IncomeTax	0.0781	0.0847	0.3238

Notes: Bolded values indicate significance at 95% confidence level. Numbers in brackets are 95% confidence intervals. Estimates obtained using the Krinsky-Robb procedure with 10,000 draws. Willingness to pay for attribute x is calculated as β_x/β_{price}. In the bottom panel, p values are reported using the complete combinatorial method of testing for differences in distributions (Poe, Giraud and Loomis 2005).

Preferences and WTP over program details also vary by treatment. Respondents in both HP and CT treatments prefer PES and regulation programs relative to the omitted program (fertilizer taxes that increase food prices). In the HP treatment, this preference is significant and generates relatively large WTP values. By contrast, while the CT treatment generates positive WTP estimates for PES and regulation programs, these values are modest relative to the HP treatment and only one program type (PES funded with sales taxes) has positive WTP values for the entire 95% confidence interval. Like the CT group, the control group exhibits a statistically significant positive preference for only one program type (PES funded with sales taxes) relative to fertilizer taxes, but WTP estimates for this program are more in line with the HP treatment than the CT treatment.

Table 5 and Figure 2 summarize how price sensitivity changes over time and whether temporal changes in price sensitivity vary by treatment group. This analysis allows us to test whether the effect of our hypothetical bias mitigation techniques persists throughout the entire set of choice exercises or fades over time. Table 5 displays regression results for both a baseline estimation that includes price-exercise interactions (Column I) and an estimation that includes three-way price-treatment-exercise interactions (Column II). Figure 2 displays a comparison of price sensitivity by both treatment and exercise. In corroboration with our previous models, the CT treatment produces larger estimates of price sensitivity than the control and HP treatments, and there is no discernible difference between honesty and neutral priming.[9] Importantly, as respondents progress to later exercises, the gap in price sensitivity between CT and the other groups narrows. This suggests that the ability for cheap talk to mitigate hypothetical bias fades

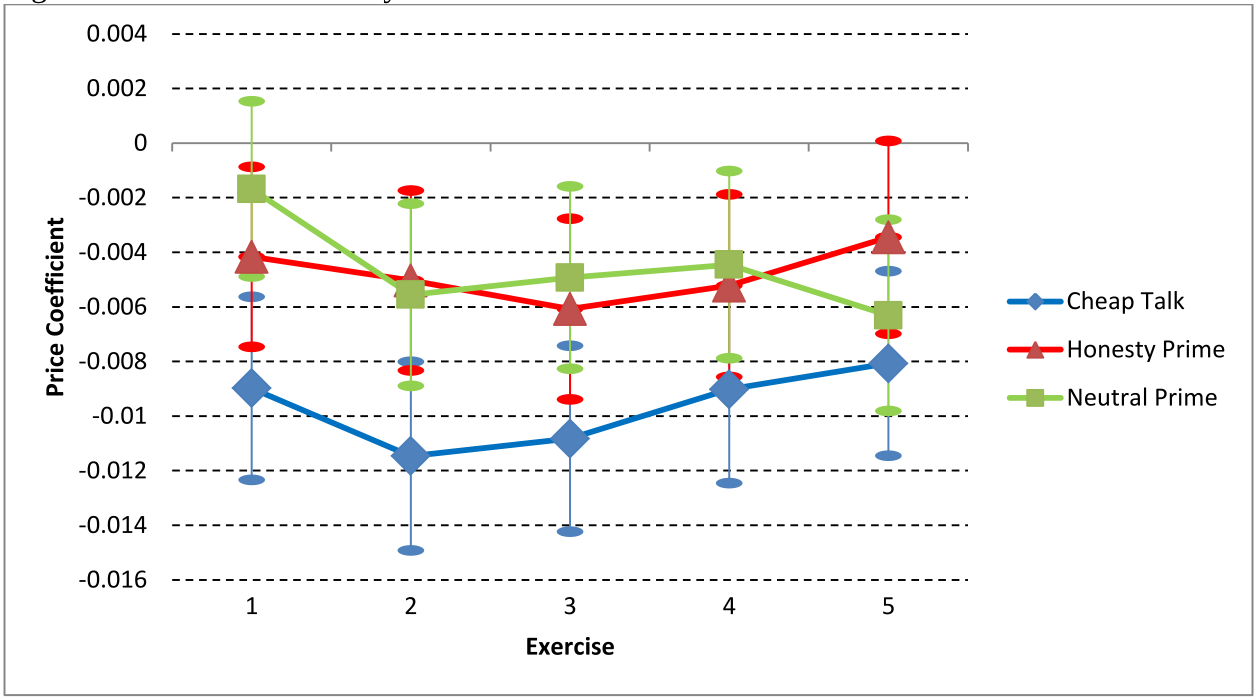
[9] Note that greater levels of price sensitivity correspond to larger (in absolute value) negative numbers, so downward movement in Figure 2 corresponds with greater price sensitivity.

Table 5: Models that Examine Differential Treatment Effects by Choice Exercise

Variable		**I**	**II**
Price	**Ex1**	-0.0048** (<0.005)	-0.0017 (0.313)
	Ex2	-0.0072** (<0.005)	-0.0055** (<0.005)
	Ex3	-0.0072** (<0.005)	-0.0049** (<0.005)
	Ex4	-0.0061** (<0.005)	-0.0044* (0.011)
	Ex5	-0.0059** (<0.005)	-0.0063** (<0.005)
Price*CT	**Ex1**		-0.0073** (<0.005)
	Ex2		-0.0059** (0.010)
	Ex3		-0.0059** (0.009)
	Ex4		-0.0046 (0.052)
	Ex5		-0.0017 (0.463)
Price*HP	**Ex1**		-0.0025 (0.240)
	Ex2		-0.0005 (0.811)
	Ex3		-0.0012 (0.604)
	Ex4		-0.0007 (0.759)
	Ex5		0.0029 (0.217)
Effect, Detail, Program A and Program B Controls?		Yes	Yes

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 18,100 observations come from 1,209 respondents, with robust standard errors clustered by respondent. *P*-values reported in parentheses.

Figure 2: Price Coefficients by Treatment and Exercise



Notes: Bars represent 95% confidence intervals around each point estimate.

during later rounds of choices. Indeed, we find the hypothetical bias mitigation effect of cheap talk scripts may decline relatively quickly.

Table 6 adds statistical rigor to our findings from Figure 2. Where appropriate, we use *t* tests (Control vs. HP and Control vs. CT) and Wald tests (CT vs. HP) and strongly reject the null of equal price sensitivity between CT and the control group for exercises 1-3, but we can no longer reject the null of equality for exercises 4 and 5. The same pattern holds when CT and HP are compared: we reject the null of equality for exercises 1-3 but not for exercises 4 and 5.

To test our conclusions for robustness, we use data from what is, to our knowledge, the only other research comparing cheap talk and honesty priming treatments to a neutral priming

Table 6: Tests for Equality of Price Coefficient: P Values

	Ex1	Ex2	Ex3	Ex4	Ex5
CT=HP	0.028*	< 0.005**	0.038*	0.099	0.052
CT=Control	< 0.005**	0.010**	0.009**	0.052	0.463
HP=Control	0.240	0.811	0.604	0.759	0.217

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. The null hypothesis is equality of the price coefficient for both groups. Comparisons of treatments with the control use t-tests, while comparisons of the two treatments use Wald Tests.

control (DGN, 2013). While DNG compare multiple additional treatments (7 total comparison groups), we restrict our focus to the three that most closely resemble the groups in our data.[10]

There are several notable differences in our data and the DGN data. The choice experiment in DGN compared different almond products with a no buy option, while our choice experiment compared different programs for reducing nutrient pollution in Lake Erie with the status quo. DGN also use a different form of honesty priming (scrambled sentences vs. synonym matching) and engage subjects in more repetitions of the priming task (24 vs. 6). Each choice exercise was administered using different media, with DGN using a face-to-face contact with subjects in the context of a lab experiment and our subjects completing an online survey. There are also significant differences in choice set design. Subjects in our study complete five choice sets, and the order of choice set presentation is randomized. DGN present 16 choice sets to each subject and do not randomize choice set order. The absence of randomization makes it impossible to separate the effect of choice exercise order from the design (specifically, attribute levels) of a specific choice exercise. To deal with this and smooth out any exercise-by-exercise variation, we combine exercises in groups of 4 when using DGN data.[11]

[10] We use data from DGN's hypothetical neutral prime, hypothetical honesty prime, and hypothetical cheap talk groups and exclude the hypothetical baseline, real baseline, real neutral prime and real honesty prime groups.

[11] We also include an analysis of the DNG data where exercises are aggregated in groups of two instead of four. Our findings are robust to the different aggregations.

Table 7: Differential Treatment Effects by Choice Exercise using data from de-Magistris, Gracia and Nayga (2013)

Variable		Coefficient
Price	**Ex1-4**	-1.2147** (<0.005)
	Ex5-8	-1.4795** (<0.005)
	Ex9-12	-1.4052** (<0.005)
	Ex13-16	-1.5351** (<0.005)
Price*CT	**Ex1-4**	-0.4037** (0.007)
	Ex5-8	-0.1990 (0.246)
	Ex9-12	-0.1834 (0.238)
	Ex13-16	-0.2890 (0.055)
Price*HP	**Ex1-4**	-0.3689** (0.010)
	Ex5-8	-0.3003* (0.048)
	Ex9-12	-0.1542 (0.309)
	Ex13-16	-0.2702 (0.051)
Controls for other Attributes and "No Buy" Option		Yes

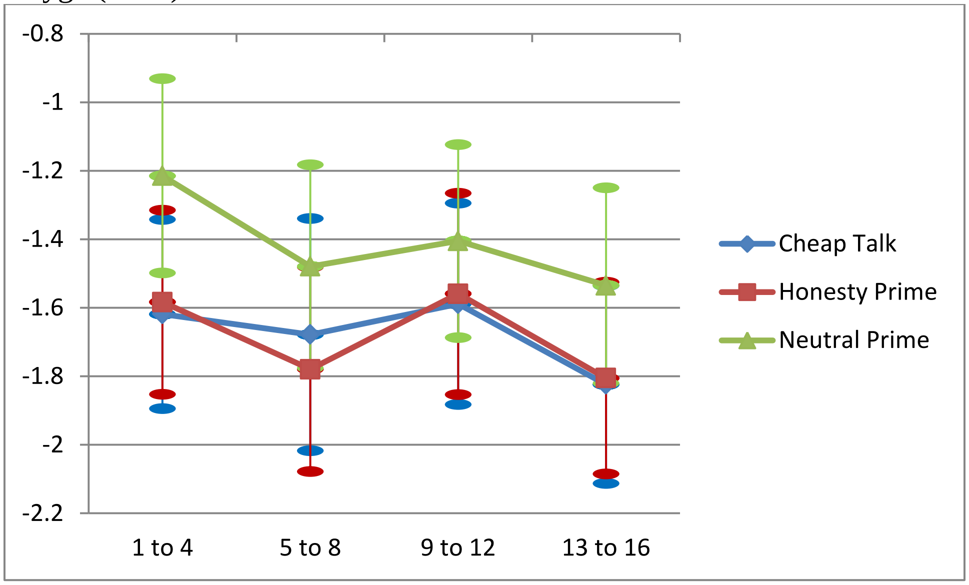
Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 7,632 observations come from 159 respondents, with robust standard errors clustered by respondent. *P*-values reported in parentheses.

Table 8: Tests of Equality P Values using data from de-Magistris, Gracia and Nayga (2013)

	Ex1-4	**Ex5-8**	**Ex9-12**	**Ex13-16**
CT=HP	0.786	0.522	0.839	0.888
CT=Control	0.007**	0.246	0.238	0.055
HP=Control	0.010**	0.048*	0.309	0.051

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. The null hypothesis is equality of the price coefficient for both groups. Comparisons of treatments with the control use *t*-tests, while comparisons of the two treatments use Wald Tests.

Figure 3: Price Coefficients by Treatment and Exercise using data from de-Magistris, Gracia and Nayga (2013)



Notes: Bars represent 95% confidence intervals around each point estimate.

The results of this analysis, in Tables 7 and 8 and Figure 3, support the finding of cheap talk erosion in our data. In the DGN data, both HP and CT increase price sensitivity. When coefficients are allowed to vary by choice exercise, the gap between control and treatment coefficients is larger in early exercises than in later ones, suggesting erosion of the treatment effect. While treatment effects are consistently negative (implying greater price sensitivity in the treatments than the control), Table 8 shows that these differences are more likely to be statistically significant in the early exercises than in the later ones. The main difference between our data and the DNG data is that, while treatment effects dissipate after the early exercises in both datasets, they appear to spike again in the last exercises of the DGN data. This spike is not statistically significant when exercises are aggregated in groups of four, but it is for CT when exercises are aggregated in groups of two (see appendix). One possible explanation for this difference across studies is differences in subject information regarding the number of choice exercises. In our sample, respondents were not aware of how many choice exercises they would perform. In DGN, subjects knew they would be making 16 choices. Subjects may choose to act differently for their final few choices, and this tendency may be magnified by the treatments.

5. Discussion and Conclusions

Our study provides further evidence that cheap talk scripts can significantly mitigate hypothetical bias in online choice exercises and suggests that honesty priming may not be as effective in online formats as it has been found in the face-to-face experimental setting of DGN. Admittedly, there are differences between our study and DGN beyond the online/face-to-face aspect of the choice experiment. As previously noted, DGN uses a different type of priming task (scrambled sentences vs. matching synonyms) and the DGN prime involves many more repetitions of the task (24 vs. 6). We believe these design differences are less likely to explain our different results than the change in survey medium. While the primes are different, both tasks have not only been demonstrated to effectively prime subjects, but have been specifically shown to prime for honesty/truthfulness in other settings.

The repetitions disparity is a plausible explanation, and indeed we find that subjects spent much more time on the cheap talk task than either priming task. However, this does not seem to explain the discrepancy. When we control for the interaction of time spent on the treatment/control task and price sensitivity, and further allow for this interaction to vary by treatment/control group, we find that the interactions of time, treatment and price are not significant, while the general effects persist (i.e. cheap talk still increases sensitivity to price and honesty priming does not).[12] This supports our conclusion that differences in treatment effects are due to the nature of the treatments rather than being a function of time spent on each task.

Our exercise-level analysis reveals a conclusion that is novel yet unsurprising: ex-ante treatments that mitigate hypothetical bias in choice experiments have the potential to fade over time. Treatments that are initially very effective did not maintain their potency, even though our study used only five exercises, which is modest compared to many designs in the literature. In an attempt to identify whether this result is an artifact of our study or indicative of a more widespread phenomenon, we use data from DGN and find the same pattern for both cheap talk and honesty priming. All treatments that mitigate price sensitivity in the aggregate (cheap talk for our data, both cheap talk and honesty priming for DGN data) exhibit the same basic pattern, regardless of whether they are overt or inconspicuous in nature.

This result has important ramifications for future choice experiment design. Choice experiments come in many forms, and a multitude of decisions can impact researchers' ability to accurately elicit preferences. Each decision is not made in a vacuum; instead the appropriateness of one choice depends on other choices. Our analysis suggests that one phenomenon, cheap talk erosion, may persist in a wide number of designs and contexts. Determining how choice experiments can be designed to eliminate cheap talk and honesty priming erosion is worthy of future consideration.

Our results suggest that past analyses of choice experiments featuring cheap talk interventions and multiple choice sets might be revisited with analyses limited to the early rounds of experimentation to further explore the robustness of our findings. If our results are

[12] *P* values for the price/CT, price/HP, price/CT/Time and price/HP/Time interactions are < 0.005, 0.850, 0.516, and 0.999, respectively.

supported by further work, then interventions that reiterate the main assertions of the cheap talk script at the mid-point of repeated rounds, or repeated priming exercises, may serve as a booster shot for hypothetical bias mitigation. Indeed, such dynamics may explain results found by Ladenburg and Olsen (2014).

References

Banzhaf, H., D. Burtraw, D. Evans, and A. Krupnick. 2006. "Valuation of Natural Resource Improvements in the Adirondacks." *Land Economics*, 82 (3): 445-464.

Bateman, I., D. Burgess, W.G. Hutchinson and D. Matthews. 2008. "Learning Design Contingent Valuation (LDCV): NOAA Guidelines, Preference Learning and Coherent Arbitrariness." *Journal of Environmental Economics and Management*, 55: 127-141.

Bech, M., T. Kjaer and J. Lauridsen. 2011. "Does the Number of Choice Sets Matter? Results from a Web Survey Applying a Discrete Choice Experiment." *Health Economics*, 20: 273-286.

Bell, J., J. Huber and W. Viscusi. 2011. "Survey Mode Effects on Valuation of Environmental Goods." *International Journal of Environmental Research and Public Health*, 8: 1222-1243.

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

Bosworth, R and L. Taylor. 2012. "Hypothetical Bias in Choice Experiments: Is Cheap Talk Effective at Eliminating Bias on the Intensive and Extensive Margins of Choice?" *The B.E. Journal of Economic Analysis and Policy*, 12(1): 1-26.

Brouwer, R., T. Dekker, J. Rolfe and J. Windle. 2010. "Choice Certainty and Consistency in Repeated Choice Experiments." *Environmental and Resource Economics*, 46: 93-109.

Canavari, M., G. Nocella and R. Scarpa. 2005. "Stated Willingness-to-pay for Organic Fruit and Pesticide Ban." *Journal of Food Products Marketing*, 11 (3): 107-134.

Carlsson, F., M. Kataria, A. Krupnick, E. Lampi, A. Lofgren, P. Qin, and T. Sterner. 2013. "The Truth, the Whole Truth, and Nothing But the Truth – A Multiple Country Test of an Oath Script." *Journal of Economic Behavior and Organization*, 89: 105-121.

Carlsson, F., P. Frykblom and C. Lagerkvist. 2005. "Using Cheap-talk as a Test of Validity in Choice Experiments." *Economics Letters*, 89 (2): 147-152.

Carlsson, F., M. Morkbak and S Olsen. 2012. "The First Time is the Hardest: A Test of Ordering Effects in Choice Experiments." *Journal of Choice Modelling*, 5(2): 19-37.

Carson, R. and T. Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics*, 37: 181-210.

Champ, P., R. Moore and R. Bishop. 2009. "A Comparison of Approaches to Mitigated Hypothetical Bias." *Agricultural and Resource Economics Review*, 38 (2): 166-180.

Chartrand, T. and J. Bargh. 1996. "Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions." *Journal of Personality and Social Psychology*, 71 (3): 464-478.

Cummings, R. and L. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Desing for the Contingent Valuation Method." *American Economic Review*, 89 (3): 649-665.

de-Magistris, T., A. Gracia and R. 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.

de-Magistris, T. and S. Pascucci. 2014. "The Effect of the Solemn Oath Script in Hypothetical Choice Experiment Survey: A Pilot Study." *Economics Letters*, 123: 252-255.

Fox, J., J. Shogren, D. Hayes and J. Kliebenstein. 1998. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics*, 80 (3): 455-465.

Haab, T. and K. McConnell. 2002. *Valuing Environmental Natural Resources: The Econometrics of Non-Market Valuation*. Northhampton, MA: Edward Elgar Publishing.

Harrison, G. and E. Rutstrom. 2008. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." *Handbook of Experimental Economics Results*, 1 (5): 752-767.

Holt, C. and S. Laury. 2005. "Risk Aversion and Incentive Effects: New Data without Order Effects." *American Economic Review*, 95 (3): 902-904.

Howard, G. 2014. "Did Money Change Them? Examining Hypothetical Bias when Eliciting Preferences for Personal and Social Benefits." Unpublished manuscript.

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

Krinsky, I. and A. Robb. 1986. "On Approximating the Statistical Properties of Elasticities." *Review of Economics and Statistics*, 68: 715-9.

Kuhfeld, W. 2010. *Marketing Research Methods in SAS*. Cary, NC: SAS Institute Inc.

Ladenburg, J., and Olsen, S. B. (forthcoming). "Augmenting Short Cheap Talk Scripts with a Repeated Opt-out Reminder in Choice Experiment Surveys." *Resource and Energy Economics*.

Ladenburg, J. and S. Olsen. 2008. "Gender-specific Starting Point Bias in Choice Experiments: Evidence from an Empirical Study." *Journal of Environmental Economics and Management*, 56: 275-285.

Lindhjem, H. and S. Navrud. 2011. "Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes." *International Review of Environmental and Resource Economics*, 5: 309-351.

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

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

Loomis, J. 2011. "What's to Know about Hypothetical Bias in Stated Preference Valuation Studies?" *Journal of Economic Surveys*, 25 (2): 363-370.

Loomis, J. 2014. "2013 WAEA Keynote Address: Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys." *Journal of Agricultural and Resource Economics*, 39 (1): 34-46.

Louviere, J., D. Hensher and J. Swait. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge: Cambridge University Press.

Ludsin, S.A., Kershner, M.W., Blocksom, K.A., Knight, R.L., and Stein, R.A. 2001. "Life after Death in Lake Erie: Nutrient Controls Drive Fish Species Richness, Rehabilitation," *Ecological Applications* 11: 731-746.

Makarewicz, J.C., and P. Bertram. 1991. "Evidence for the Restoration of the Lake Erie Ecosystem," *Bioscience* 41:216-223

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: 388-398.

Matisoff, G., and J. J. H. Ciborowski. 2005. "Lake Erie Trophic Status Collaborative Study," *Journal of Great Lakes Research* 31(Suppl. 2):1-10.

Moore, R., R. Bishop, B. Provencher and P. Champ. 2010. "Accounting for Respondent Uncertainty to Improve Willingness-to-pay Estimates." *Canadian Journal of Agricultural Economics*, 58: 381-401.

Morrison, M. and T. Brown. 2009. "Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies." *Environmental and Resource Economics*, 44: 307-326.

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.

Nielson, J. 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: 119-129.

Ohio Environmental Protection Agency. 2010. Ohio Lake Erie Phosphorus Task Force Report. Columbus, OH. 109p.

Olsen, S. 2009. "Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-market Goods." *Environmental and Resource Economics*, 44: 591-610.

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

Rasinski, K., P. Visser, M. Zagatsky and E. Rickett. 2004. "Using Implicit Goal Priming to Improve the Quality of Self-report Data." *Journal of Experimental Social Psychology*, 41 (3): 321-327.

Satterhwaite, M. 1975. "Strategy-proofness and Arrow Conditions: Existence and Correspondence Theorems for Voting Procedures and Welfare Functions." *Journal of Economic Theory*, 10: 187-217.

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

Silva, A., R. Nayga, B. Campbell and J. Park. 2012. "Can Perceived Task Complexity Influence Cheap Talk's Effectiveness in Reducing Hypothetical Bias in Stated Choice Studies?" *Applied Economics Letters*, 19: 1711-1714.

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

Vinski, M. and S. Watter. 2012. "Priming Honesty Reduces Subjective Bias in Self-report Measures of Mind Wandering." *Consciousness and Cognition*, 21: 451-455.

Whitehead, J. and T. Cherry. 2007. "Willingness to Pay for a Green Energy Program: A Comparison of Ex-ante and Ex-post Hypothetical Bias Mitigation Approaches." *Resource and Energy Economics*, 29: 247-261.

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.

Appendix

Table 9: Differential Treatment Effects by Choice Exercise using data from de-Magistris, Gracia and Nayga (2013), Groups of Two

Variable		**Coefficient**
Price	**Ex1-2**	-1.2126** (<0.005)
	Ex3-4	-1.2263** (<0.005)
	Ex5-6	-1.5030** (<0.005)
	Ex7-8	-1.4589** (<0.005)
	Ex9-10	-1.4049** (<0.005)
	Ex11-12	-1.4022** (<0.005)
	Ex13-14	-1.6041** (<0.005)
	Ex15-16	-1.4452** (<0.005)
Price*CT	**Ex1-2**	-0.4002 (0.059)
	Ex3-4	-0.4066* (0.013)
	Ex5-6	-0.2125 (0.231)
	Ex7-8	-0.1827 (0.362)
	Ex9-10	-0.1835 (0.236)
	Ex11-12	-0.1891 (0.577)
	Ex13-14	-0.1969 (0.214)
	Ex15-16	-0.4283* (0.035)

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 7,632 observations come from 159 respondents, with robust standard errors clustered by respondent.

Table 9 Continued

Price*HP	**Ex1-2**	-0.4283* (0.019)
	Ex3-4	-0.3432* (0.036)
	Ex5-6	-0.3060 (0.060)
	Ex7-8	-0.2931 (0.120)
	Ex9-10	-0.1320 (0.373)
	Ex11-12	-0.2815 (0.432)
	Ex13-14	-0.2677 (0.093)
	Ex15-16	-0.2646 (0.183)
Controls for other Attributes and "No Buy" Option		Yes

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively. 7,632 observations come from 159 respondents, with robust standard errors clustered by respondent. *P*-values reported in parentheses.

Figure 4: Price Coefficients by Treatment and Exercise using data from de-Magistris, Gracia and Nayga (2013), Groups of Two

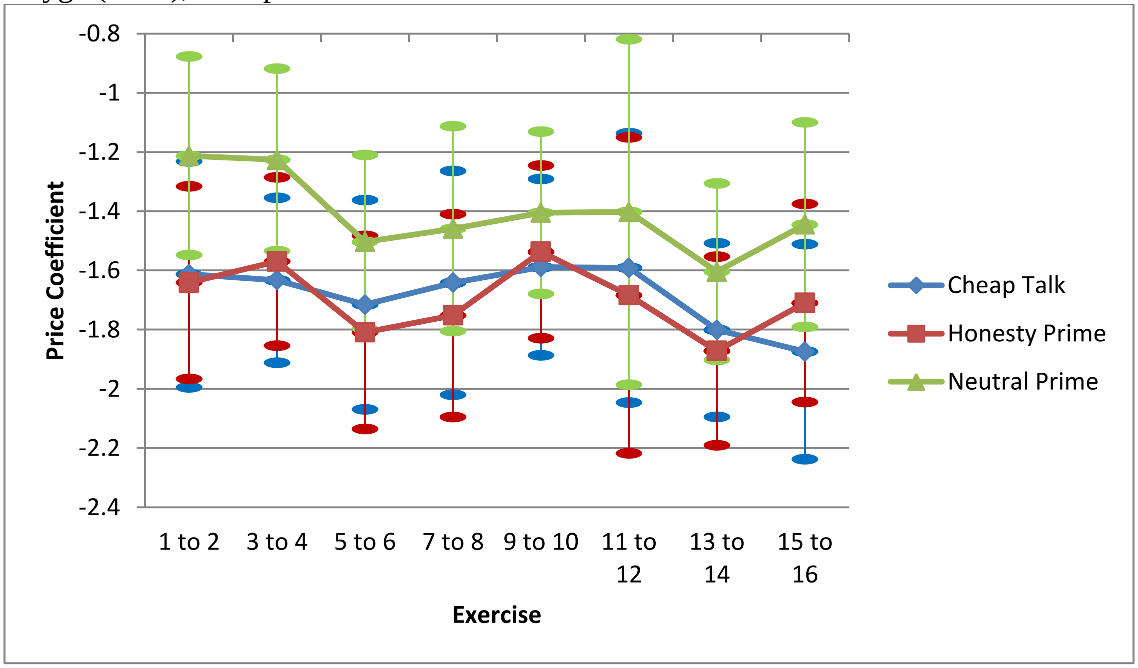


Table 10: Tests of Equality P Values using data from de-Magistris, Gracia and Nayga (2013), Groups of Two

	Ex1-2	**Ex3-4**	**Ex5-6**	**Ex7-8**	**Ex9-10**	**Ex11-12**	**Ex13-14**	**Ex15-16**
CT=HP	0.884	0.645	0.585	0.561	0.728	0.756	0.647	0.382
CT=Control	0.059	0.013*	0.231	0.362	0.236	0.577	0.214	0.035*
HP=Control	0.019*	0.036*	0.060	0.120	0.373	0.432	0.093	0.183

Notes: * and ** indicate statistical significance at 95% and 99% confidence levels, respectively.