2017

ENVIRONMENTAL VALUES, STATED PREFERENCES, AND HYPOTHETICAL BIAS

Jerrod M. Penn
University of Kentucky, jmpenn311@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-5576-7098
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.275

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Jerrod M. Penn, Student

Dr. Wuyang Hu, Major Professor

Dr. Carl Dillon, Director of Graduate Studies

ENVIRONMENTAL VALUES, STATED PREFERENCES, AND HYPOTHETICAL BIAS

DISSERTATION

A dissertation submitted in partial fulfillment of the
Requirements for the degree of Doctor of Philosophy in Agricultural Economics in
The College of Agriculture, Food and Environment
At the University of Kentucky

By

Jerrod M. Penn

Lexington Kentucky

Director: Dr. Wuyang Hu, Professor of Agricultural Economics

Lexington Kentucky

2017

ABSTRACT OF DISSERTATION

ENVIRONMENTAL VALUES, STATED PREFERENCES,
AND HYPOTHETICAL BIAS

Contingent Valuation (CV) methods are a primary tool in environmental economics to ascertain non-use or other values not observable through existing market mechanisms. Because common CV approaches typically rely on hypothetical answers from surveys in order to generate welfare estimates, these are often labelled stated preferences. Results from stated preference methods often diverge from those obtained when actual preference or behavior are involved. This divergence is commonly known as Hypothetical Bias (HB). This dissertation addresses HB as it applies to environmental applications. To begin, a meta-analysis using a sample of studies many times larger than previous works was performed. Its results identify which study protocols exacerbate HB, and which may mitigate it. Furthermore, the meta-analysis establishes the efficacy of some popular techniques to mitigate HB. The second essay focuses on understanding and addressing two important topics to environmental economics, distance decay and charismatic species conservation. These effects have not been investigated with respect to HB. We implement a field survey of monarch and viceroy butterfly conservation, creating survey treatment conditions involving both real payment and hypothetical scenarios in order to establish the extent of HB. The key finding is that while HB is present for both butterflies, HB in distance decay exists for monarchs. There is also additional HB for monarchs compared to viceroys, which we attribute to the former's charisma. The final endeavor studies the usefulness of consequentiality, a relatively new tactic to reduce HB. Consequentiality is the degree to which respondents believe their answers may affect policy outcomes. Relying on the monarch field survey, we find that using a technique known as ex ante consequentiality may exacerbate HB. Another approach known as ex post consequentiality is more effective at reducing the extent of HB in the data. Lastly, some elements of the studies' results showcase that HB is not always present and can also explain some of the mixed results found on the efficacy of HB mitigating methods reported in previous studies.

KEYWORDS: Non-market valuation, Choice Experiment, Hypothetical Bias, Monarch, Viceroy, Consequentiality

Jerrod M. Penn

June 20, 2017

ENVIRONMENTAL VALUES, STATED PREFERENCES, AND HYPOTHETICAL BIAS

By

Jerrod M. Penn

Wuyang Hu
Director of Dissertation

Carl Dillon
Director of Graduate Studies

June 20, 2017

THIS WORK IS DEDICATED TO THOSE WHO HAVE BELIEVED, INVESTED, AND BEEN PATIENT WITH ME. SO SHALL I SERVE TOO.

ACKNOWLEDGMENTS

Many have acknowledged that we are the product of our environment. I am no different, able to succeed thanks to the support of many people. First, I must thank Dr. Wuyang Hu, my advisor and friend. He was the single most important reason I chose UK and the single most important reason for my success. I am also indebted to our Chair, Leigh Maynard. His support enabled the success of many AEC students.To my AEC family, thank you for your support and encouragement, especially overcoming hurdles to implement a number of ambitious ideas, and memories of time shared together. Thank you to Mike Lorton and LFUCG Parks & Recreation for enabling me to complete the survey included here. Thank you to my dissertation committee, Dr. Craig Landry from the University of Georgia, Dr. Glenn Blomquist, Dr. Jill Stowe, and Dr. Mark Williams from Horticulture. Their guidance and refinement has greatly improved this work. To the UK Student Sustainability Coucil and Shane Tedder, thank you for supporting this and a number of other projects. Serving UK Sustainability has been one of the most rewarding parts of my time at UK. Thank you to my students for enduring my continuous challenge to represent yourselves, our department, and our university to the utmost. You have been a defining part of my time as a UK graduate student. To these many people, I am compelled to continue to work diligently and on behalf of others in order to honor your investment in me.

Lastly, I must recognize the importance of my family. To Hannah, thank you for patience and willingness to endure my many shenanigans. To my parents, Sandra and Paul Penn. Thank you for giving me the chance to grow up the sticks where I learned to love nature and embrace hard work. You instilled in me the goal to serve others and my choices every day reflect my desire to make you proud. JC

TABLE OF CONTENTS

Table of Tables

Table of Figures

Chapter 1 INTRODUCTION

1.1 General Background

Humans continuously change their environments. These changes take many forms, sometimes pre-meditated such as large-scale highway or building construction projects or even as simple as installing a bird feeder in a backyard. Other times, the change to the environment is unplanned, such as the historic Exxon-Valdez and Deepwater Horizon oil spills to accidentally transferring emerald ash borer from moving firewood short distances. Each of the examples can create a series of costs and benefits, values that affect the well-being of people directly and indirectly. In order to create better policy design and improve societal outcomes, measuring and quantifying each component is essential.

Some changes in value are relatively straightforward to measure, such as the quantity of damaged timber from emerald ash borer. Others are more subtle, such as the lost value of people whose well-being has been diminished knowing that large proportion of forests in the Eastern United States are now composed of dead ash trees, or alternatively, the values of degraded ecosystems or disappeared animal or plant species. Measuring these latter examples is difficult in that there is no corresponding price or market to identify changes. This is a primary reason why Contingent Valuation (CV) methods are used.

As many practitioners are well aware, they espoused the usefulness of CV's is most useful to investigate non-use values, those that are not readily measurable with functioning markets or prices, nor indirectly observed through other decisions, known as the revealed preference methods. Non-use value can

represent a significant proportion of total value, so ignoring such values can lead to misinformed welfare estimates (i.e. measures of economic value) and policy decisions. As such, the capability of stated preference methods to estimate non-use value is also one of its greatest sources of doubt because asking direct questions about hypothetical scenarios often for unfamiliar goods means values can be influenced with relative ease versus other approaches. The usefulness and limitations of CV received notoriety in the assessment by the NOAA Blue Ribbon Panel (Arrow, et al., 1993) . Given these weaknesses, the key is to develop methods for more precise welfare estimates and understand when and how much various factors come into play.

1.2 Objectives and Structure

The purpose of this dissertation is to augment our understanding of stated preference methods with respect to environmental applications, with primary focus on understanding and mitigating Hypothetical Bias (HB). HB is the often-documented outcome that economic values, particularly welfare estimates from hypothetical value elicitations are different and larger than their counterparts when real transaction of money is involved when measuring these values. In this dissertation, this is first accomplished in Chapter 2 with an updated meta-analysis comparing welfare estimates from studies that implement and compare both purely hypothetical value elicitation methods as well as real elicitation methods in which actual payment was required in at least some situations. The welfare estimates are typically Willingness to Pay (WTP), or in a small number of

cases, Willingness to Accept (WTA). Compared to previous similar works using meta-analysis (List, 2001, Murphy, et al., 2005), the dataset in this chapter is roughly ten times larger in the number of observations and over three times larger with respect to the number of studies included. The meta-regression includes many of the same variables from these previous analyses, but because of the sample size and inclusion of new studies, provides opportunity for a more comprehensive analysis. For instance, the meta-regression included indicators for whether Cheap Talk, certainty follow-up question, consequentiality, or some other HB mitigation strategy was used in each of the research included in our dataset. Utilizing such an encompassing dataset from a broad set of applications can help elucidate the circumstances when HB may or may not occur.

The second essay (Chapter 3) focuses on understanding a specific phenomenon of HB with respect to distance decay and with charismatic species through a valuation of butterfly conservation. A species or environmental site is subject to distance decay if its total economic value decreases as the physical distance from species or environmental site increases. Failing to account for non-use values of populations far removed means economic value is understated, but may be overstated if the value is assumed to be equal to those closest to the resource. For example, many individuals may financially support giant panda conservation without any intent to visit China. Issues of distance and value have been recognized for decades (Sutherland and Walsh, 1985) but is still unsettled (León, et al., 2016, Schaafsma, et al., 2012), especially with the recent analysis of spatial patchiness and hotspots (Johnston and Ramachandran, 2014). No

studies have examined the relationship between WTP and distance decay and included an elicitation mechanism involving actual payment, and therefore HB with respect to spatial decay has not been appropriately studied. We investigate the presence of HB in a study of spatial decay using an application of butterflies.

In a previous meta-analysis of conservation for predominately threatened, endangered, and rare species by Loomis and White (1996), the authors found WTP values approaching $1000, a number later noted by Brown and Shogren (1998) as being "suspiciously high" such that "less than 2% of all threatened and endangered species represented 1% of the 1995 US GNP."[1] This may suggest that HB may be exacerbated by, or a proportion of it explained by, valuing charismatic species. We study this by comparing monarch and viceroy butterflies in this chapter of the dissertation.

Monarchs (*Danaus plexippus*) exhibit a bright orange color pattern, making them one of the most well-known butterflies in the United States. The monarch population has recently declined precipitously to the extent that it is currently under status review for inclusion on the endangered species list (US Fish and Wildlife Service, 2016). Additionally, its conservation is one of the three primary goals in the Obama Administration's "National Strategy to Promote the Health of Honey Bees and Other Pollinators" (Pollinator Health Task Force, 2015). As such, monarchs are *potentially* endangered and also well-known, the two criteria of our definition of a charismatic species. With this definition, the

[1] After annuitizing one-time lump-sum payments, the estimates come to a more reasonable $409 per year in 1995 dollars, or $644 in 2016 dollars.

appropriate comparison is to a non-charismatic species, ideally one that is in all ways similar to the charismatic species except that it is relatively unfamiliar to people as well as not a threatened or endangered species. This juxtaposition, especially with mammals, the typical charismatic species, is usually not available. The monarch butterfly does in fact have a doppelgänger, namely the viceroy butterfly. With respect to visual characteristics and habitat, the two are extremely similar and the viceroy does not have a compromised conservation status.

We implement a field survey in Lexington, KY in the summer of 2016 for donations to promote conservation of monarch and viceroy butterflies. This field survey collects responses in situations involving both hypothetical and real payment in order to generate Hypothetical and Real WTP. This field experiment contributes to the literature in two ways: 1) Because previous studies of distance decay have not included a real payment treatment group, we contribute by testing if and to what extent distance decay is prone to hypothetical bias. 2) We can establish WTP for various non-use values related to monarch and viceroy butterflies. By using a split sample design of the two butterflies, which are visually nearly identical, we can further uncover the value specific to the Monarch butterfly, which is a prominent insect among the public and governmental attention due to the rapid decline in its population in the past two decades. We hypothesize that Monarchs have greater WTP, and because the two species are mimics of each other, the difference in values between the two represent the distinct value of the Monarch species as an emblematic and potentially endangered species.

Chapter 4 details the third essay, which also relies on the field survey to study HB mitigation strategies, specifically the effect of consequentiality treatments to respondents' WTP. Recent literature has emphasized the importance of evaluating participants' perception of consequentiality, or the extent to which they believe their individual answers in the survey as well as the collective results of the survey can affect broader outcomes. Theory dictates that incentive-compatible (i.e. truthful) responses are only possible if the respondent believes their answer can affect outcomes, otherwise known as policy consequentiality. Those who do not (i.e. inconsequential respondents) should be excluded from analysis. Studies of policy consequentiality usually consider consequentiality treatment was implemented prior to valuation (ex ante) or after (ex post) valuation. By ex ante, we refer to the practice of subjecting respondents to a policy consequentiality treatment during the survey. By ex post, we refer to the method of adjusting the data/respondents analyzed based on responses collected in the survey after the valuation elicitation. As suggested, if respondents do not pass a series of thresholds to signal that their responses were based on sufficient perceived consequentiality, their responses are excluded from analysis. WTP measures can then be calculated and compared across different treatments. The key contribution of this study is that a real-payment treatment is included. As far as we know, similar studies have only studied effects of consequentiality in a purely hypothetical setting where no real payment was involved. With real- and hypothetical-payment treatments both assessed, we can establish a true measure of HB and address 1) the effects of

ex ante and ex post consequentiality on HB, and 2) compare the effect of ex ante consequentiality to CT on HB.

From a broader perspective, these papers examine stated preferences, how respondents answer in surveys or other hypothetical situations in which respondents may not carefully consider the consequences of their choices. As such, they are free to answer based on other elements beyond neo-classical welfare maximization. The aim of this dissertation is to understand particular elements of why this happens for environmental and conservation efforts, and mitigate these outside factors. By doing so, practitioners of such methods can more accurately describe how people interacting with or changing their environments, both significantly and immaterially, can affect the well-being of themselves and others. Further, we show how some pitfalls may be avoided in the process of generating welfare estimates that may be more acceptable by economists and policymakers. Chapter 5 summarizes the collective findings and provides some discussion of potential implications.

Chapter 2 UNDERSTANDING HYPOTHETICAL BIAS: AN ENHANCED META-ANALYSIS

2.1 Abstract

The presence of hypothetical bias (HB) associated with stated preference methods has garnered frequent attention in environmental economics, marketing studies and related literature. This study conducts an updated meta-analysis using a dataset much broader than previous meta-analyses allowing the inclusion of several important factors that have not been investigated before. These include relatively recent willingness to pay elicitation methods such as choice experiments and the Turnbull lower bound estimator. Newly emerged HB reduction techniques such as consequentiality and certainty follow up treatments are also included. For explanatory variables that have been examined in previous studies, we report inconsistent findings. New variables, such as choice experiment, consequentiality and certainty follow up all significantly contributed to explaining the magnitude of HB. These results help further explain HB's presence and its amelioration in future research endeavors.

2.2 Introduction

Stated preference approaches have become a staple technique to understand consumers' values in many areas of economics, including food (Meas, et al., 2015), the environment (Juutinen, et al., 2011), and health (Cameron and DeShazo, 2013). Yet, recognition of hypothetical bias (HB), the difference in stated values versus real values, has existed as long as the

approach. List and Gallet (2001) (LG), Little and Berrens (2004) (LB), Murphy, et al. (2005) (MASW), and most recently, Little, et al. (2012) (LBB) systematically documented the persistence of HB via meta-analysis.

Each previous meta-analysis examined multiple studies that implemented and compared a hypothetical valuation, in which the respondent is not responsible to the financial commitments they may have stated, to a real valuation, in which stated financial commitments have a non-zero probability of being binding. These meta-analyses also documented details and processes of each study. Commonly studied variables included the elicitation technique used, the type of good, the use of student participants, etc. Many of these variables were found to be significant predictors of the presence of HB. Given the explosion of the more recent literature, especially studies using choice experiment and experimental methods, on the issues related to HB mitigation, there is a need to update the meta-analysis.

This study utilizes a greatly expanded dataset to enable a more comprehensive and refined examination on various experimental protocols' effect on the magnitude of HB. Comprehensiveness is improved by using a Turnbull lower bound of Willingness to Pay in order to include previously overlooked articles, which only reported the proportion of respondents who agreed to a given price level in the elicitations. Furthermore, relatively few choice experiment results existed in the literature at the time of LG, LB and MASW, and even fewer were included in their meta-analyses. While the work of Little, Broadbent and Berrens (2012) is recent, they rely on an indicator variable for the presence of HB

rather than the actual magnitude of HB. Murphy et al. 2005 is the most recent meta-analysis to utilize the magnitude of HB, but the most recent work in their sample was published in 2003. Consequently, this paper uses a much larger meta-analysis dataset to test if previously examined factors and unexplored characteristics significantly determines the magnitude of HB.

The remainder of the paper is organized as follows. A brief synopsis outlines the history of HB including explanatory theories and mitigation techniques utilized, continued with an outline of the article (data) collection process and minimum requirements for inclusion in the meta-analysis. The variables used, the econometric methods, and variants of variable and model specifications are then described. Results are presented, followed by discussion and implications.

2.3 Background

While documentation of HB occurred extensively beforehand, attention and criticism increased substantially with the Exxon-Valdez Oil Spill and the subsequent NOAA Blue Ribbon Panel report (Arrow, et al., 1993). Although HB is not always present, especially in induced-value experiments (e.g. Taylor, et al. (2001), Vossler and McKee (2006), and Mitani and Flores (2007)), work to explain and mitigate its presence have flourished.

A number of explanations of why HB continues to persist in stated preference approaches have been explored.[2] Some have used a psychology framework, such as Mitchell and Carson's (1989) assertion that stated WTP is intention that is affected by correspondence (the correlation between intention and attitudes to actual behavior), proximity (the degree to which a hypothetical decision mimics a real decision such as using voting intention instead of political attitude to predict an election) and familiarity (the level of cognizance and knowledge of the behavior under consideration). Ajzen, et al. (2004) examined a number of reasons from a social psychology framework based on the theory of planned behavior. They argued that planned behavior is affected by intention, which is based on attitudes, subjective norms, and perceived behavioral control. Vlaev (2012) tested the effect of cognitive biases on both hypothetical and real choices during social decision-making. Neuroscience has begun to study brain activity to understand differences in stated and real decisions of respondents (Kang, et al., 2011). Use of psychology is even more important to some mitigation techniques such as honesty priming (de-Magistris, et al., 2013) and elicitation under oath (Jacquemet, et al., 2013).

Other evidence suggests respondents' socioeconomic factors affect the magnitude of HB. HB is found to be more prevalent among males versus females (Brown and Taylor, 2000, Mitani and Flores, 2007). Similarly, men and women

[2] The vast majority of the literature points to an upward bias such that hypothetical values exceed real values (Schulze et al., 1981), rather than random bias (Mitchell and Carson, 1989).

may respond differently to HB mitigation techniques (Ladenburg, et al., 2011, Mahieu, 2010). Based on three separate datasets of primarily South Korean residents' responses, Mjelde, et al. (2012) developed a bias ratio, the ratio of respondents who change their choice under hypothetical versus real payment settings, to explain and mitigate discrepancies in hypothetical and real values. They found that neither income nor gender influences the bias ratio, but education and age do.

As much as HB has persisted throughout stated preference valuation studies, so have the efforts to eliminate and mitigate its presence using various techniques, as is well documented in Loomis (2011) and Loomis (2014). An initial concern that is still under study is the incentive compatibility of the elicitation format (see Carson and Groves (2007) and Carson, et al. (2014)). Critique of early works' problems of apparent free-riding and strategic overbidding led to more refined incentive structures and provision mechanisms. Adding a minimum provision point mechanism such that contributions are refunded if a minimum is not met[3] significantly improves the alignment of hypothetical and real WTP (Poe, et al., 2002). Similarly, the level of certainty that payment will happen or that the good will be provided affects stated values (Mitani and Flores, 2014).

Others have studied how elicitation format affects WTP and HB (Cameron, et al., 2002, Hoehn and Randall, 1987, Vossler and McKee, 2006). For example, comparisons show dichotomous choice typically overstates WTP relative to

[3] Akin to crowd-sourced funding where individuals commit to a financial pledge but only actually pay if a minimum dollar amount is met.

open-ended (Balistreri, et al., 2001, Schulze, et al., 1996) and payment card formats (Ready, et al., 1996, Welsh and Poe, 1998). Others have compared dichotomous choice to choice experiments but found fewer differences in WTP estimates across these methods (Christie and Azevedo, 2009, Loomis and Santiago, 2013). However, even real, non-hypothetical WTP settings may not generate consistent results across elicitation formats (Gracia, et al., 2011).

One of the earliest *ex ante* mitigation techniques introduced and still frequently used is cheap talk, made popular by Cummings and Taylor (1999), which informs respondents of HB and/or reminds them to answer the hypothetical valuation question as if it were a real and binding purchase. While its efficacy is mixed, cheap talk continues to receive considerable attention (Mahieu, et al., 2012, Silva, et al., 2011).

Use of a certainty follow-up question is among the most popular *ex post* corrections in stated preference valuations. This method provides a second question immediately following the valuation query, asking how confident the respondent is of their previous response. Early studies of its effectiveness were done by Champ, et al. (1997) and Blumenschein, et al. (1998) and more recent, elaborate tests were conducted by Blomquist, et al. (2009) and Ready, et al. (2010).

Recently, attention and study of consequentiality has grown (Carson and Groves, 2007, Interis and Petrolia, 2014). An important distinction of consequentiality is its theoretical justification for affecting Hypothetical Bias (Carson, et al., 2014), which cannot be said of other common techniques such as

Cheap Talk and certainty follow-ups. Consequentiality can be applied both as an ex ante, exogenous intervention (much in the same way as Cheap Talk) or an ex post, endogenous answer from the respondent (similar to a certainty follow-up question), both means of adjusting for the respondent's perceived consequentiality of the valuation component. For example, Bulte, et al. (2005) found an ex ante consequentiality script significantly lowered WTP, Interis and Petrolia (2014) found that ex post perceptions of consequentiality affect welfare estimates, and Herriges, et al. (2010) implemented both. Furthermore, consequentiality can be decomposed into payment consequentiality, policy consequentiality, or both (i.e. strong consequentiality), which accounts for the respondent's belief that their survey answers affect real policy making. Mitani and Flores (2014) and Vossler, et al. (2012) compared the aspects of both forms of consequentiality. Finally, Carson and Groves (2007) suggested that the role consequentiality treatment may play depends on whether a public or a private good/service is in discussion. Given the requirements for inclusion, too few studies exist to allow a credible examination of these additional aspects of consequentiality in this current meta-analysis but an investigation at a future time is warranted.

Others mitigation techniques exist that have received less attention. Early techniques include budget and substitute good reminders (Loomis, et al., 1994, Neill, et al., 1994), dissonance minimization, which gives respondents a chance to voice support *without* a financial commitment (Blamey, et al., 1999, Loomis, et al., 1999, Morrison and Brown, 2009), and ex-post calibration, which adjusts

hypothetical WTP based on other information after the data are collected (Fox, et al., 1998, List, et al., 1998). Development of correction methods has continued, with recent approaches including the solemn oath (Jacquemet, et al., 2013), Bayesian truth serum (Barrage and Lee, 2010, Weaver and Prelec, 2013), honest priming (de-Magistris, et al., 2013), and religious priming (Stachtiaris, et al., 2011). The effort by so many to 'solve HB' serves as evidence of its importance and the usefulness of this analysis.

2.4 Data Collection

The credibility of a meta-analysis relies on the articles used as well as careful scrutiny per article. To the best of our ability, we follow the protocols described in Stanley, et al. (2013), such as literature search methods, coding, and variable consideration, described hereafter. In order to be included in the analysis, the study *must* have implemented a real treatment. This precludes those who only use stated preference methods, especially those that focus on HB mitigation techniques such as Carlsson, et al. (2005), Bedate, et al. (2009), or Carlsson, et al. (2013). Likewise, articles that did not include a hypothetical treatment were excluded; for instance, Maynard, et al. (2004), Alfnes, et al. (2006), Corrigan, et al. (2009) or Michaud, et al. (2013). Articles that implicitly reveal WTP via travel cost or hedonic methods to provide a real WTP estimate are excluded. This follows the norm established by the previous meta-analyses who also exclude articles such as Adamowicz, et al. (1994), Fix and Loomis (1998), among others. Carson, et al. (1996) and Shrestha and Loomis (2001, 2003) cover these types of works extensively. Similarly, some articles evaluated

HB (e.g. Birol, Smale & Yorobe, 2012 and Boyle et al., 1996) by comparing a hypothetical treatment to responses from a revealed preference group, defined as those who had previous experience using the good. We exclude these articles since the revealed preference group did not actually have a binding financial commitment. To increase the number of studies included, we allow non-peer-reviewed articles such as Boyce, et al. (1989), Kimenju, et al. (2006), and Jianjun (2008), as was also done in MASW and LBB.

Identifying articles for this meta-analysis came from a number of sources. The first approach relied on a search in EconLit of "Hypothetical Bias," which yielded 123 published articles as of January 2015. Of these, 57 articles were relevant and incorporated into the dataset. The second approach had two steps. The first step identified articles from previous meta-analyses including MASW and LBB. In the second step, we inspected the literature reviews and citations of the first step's meta-analysis articles, checking for other related work that met the necessary criteria. The second approach also used Google Scholar to search for more recent studies that cited articles from the first approach. This process added an additional 75 articles, for a total of 132 studies in the meta-analysis. This includes 24 of 29 articles from LG,[4] all articles from MASW (28), and 85 of 96 of articles from LBB. Every article considered in this meta-analysis was

[4] The excluded articles are Bishop and Heberlein, 1990; Boyce et al., 1992; Coursey et al. 1987; McClelland et al 1993; and Navrud 1992. Explanations are provided in Appendix A in the supplementary appendix online.

downloaded, individually inspected, and highlighted for relevant passages of information pertaining to the variables of this study.[5]

Accurately coding the characteristics of each study is equally important to the credibility of a meta-analysis. To ensure an accurate characterization of each article, this study's coding was compared to the meta-analysis datasets of MASW and LBB, and from other studies' tables. [6] Two individuals were responsible for coding each article, with a third for random spot-checking. In certain circumstances, assigning a value to a variable was unclear from the study, so author discretion was used to code specific variables. Appendix A (in a separate file) documents coding choices and justification per study.

In specifying a dependent variable, measuring the *magnitude* of HB is crucial, which necessitates point estimates such as mean and median WTP in the real and hypothetical treatments. As in LG and MASW, we use a Calibration Factor (CF), the ratio of hypothetical WTP to real WTP as the dependent variable. We included all available CFs per study regardless of whether the underlying WTP values were statistically significant. The meta-analyses by LB

[5]A pdf of each article (with relevant portions of text highlighted) as well as a complete list of all the studies included in the meta-analysis is available from the authors upon request.

[6] They are Brown et al (1996), Foster et al. (1997), Byrnes et al. (1999), List and Gallet (2001), List and Shogren (2002), List (2003), Burton et al. (2007), Harrison and Rutstrom (2008), Broadbent et al. (2010), Silva et al. (2011), and Fifer et al. (2014). We also used these above articles' literature review tables to find other potential articles.

and LBB used an indicator variable for the presence or absence of HB per study, but this method may not be well-suited to capture the potentially subtle effects of experimental protocols on HB. At the same time, these meta-analyses benefit from modelling an indicator variable since they can include those studies that test for the presence of HB without relying on amounts of WTP, such as Onwujekwe, et al. (2005) and Barrage and Lee (2010).

To overcome the shortcoming of including only a limited number of studies in the meta-analysis and still maintain a cardinal dependent variable of HB, we infer a non-parametric point estimate of WTP using the Turnbull lower bound estimate as described by Haab and McConnell (2002, pg. 72-78). This enables us to incorporate more studies into the sample that did not specifically provide WTP, but did report proportions of responses that indicated yes to the valuation question from dichotomous choice or referenda elicitations, such as Landry and List (2007) and Ehmke, et al. (2008). Using this technique allowed for 126 additional observations from 33 studies in the full sample of 132 articles used in this study's analysis. The formula for the Turnbull lower bound estimate of WTP in a single price dichotomous choice or referenda elicitation is simply the proportion of yes responses multiplied by the single price. Of the 33 studies, 21 used a single price. Refer to Haab and McConnell (2002) for calculating the Turnbull lower bound of WTP in a multi-price setting.

While not ideal, using this approach provides a consistent estimate of the lower bound of expected WTP, and because it is applied to both hypothetical and real WTP, no additional bias will be introduced by this method to the two WTP

elicitation methods. Additionally, we inspected three studies[7] that provided results of both the proportion of positive responses and mean WTP estimates for each treatment in their study. The Turnbull lower bound WTP Calibration Factor deviated from the Mean WTP calibration factor by an average of 18.1%, regardless of whether a HB mitigation technique was implemented. We view this as evidence to support incorporating studies based on Turnbull WTP.

In addition to augmenting the number of articles, we reexamined and expanded observations from the articles of MASW and LG. For example, MASW consolidate multiple WTP results from multiple elicitation methods into a single observation/row for Cameron, et al. (2002), while we include 10 rows for the same study. Rather than including all observations, LG had three models that used either minimum, median or maximum CF, whereas when appropriate, we include all observations in the same dataset. Additionally, MASW excluded observations that implemented a different elicitation method in the hypothetical and real valuations, whereas we include such observations and control for the disparity accordingly.

2.4.1 Variables

[7] For Champ et al. (2009), the Turnbull calibration factors are between 10.1% and 24.5% different than the original ratios. Similarly, for Johanesson et al. (1998), the difference in the Turnbull and original WTP calibration factors are between 2.6% and 18.1%. Lastly, Turnbull estimates were between .01% and 36.8% of Blomquist et al.'s (2009) inferred calibration factors.

This study relied on the previous work of LG, LB, MASW, and LBB as a basis for many variables. **Table 2-1** outlines models from previous meta-analyses, such as the dependent variable specified, the number of studies, and results of select variables' effect on HB (a description of the variables is provided in **Table 2-2**). For example, both MASW and LBB found that student samples significantly ("SS") increased HB ("More HB"). As a second example, the use of choice experiments was associated with a lower prevalence of HB compared to dichotomous choice ("Less HB"), but was not statistically significant ("Not SS").

As mentioned earlier, the dependent variable and unit of observation is a "calibration factor" (the ratio of hypothetical WTP to real WTP), as in Foster, et al. (1997) and List and Shogren (2002). CFs maintain cardinal value, an important feature to distinguish potentially subtle differences across study characteristics. CFs are unitless, and therefore are comparable across studies, regardless of time, currency, or country. Additionally, our study can take advantage of the many recent choice experiment studies that generate multiple hypothetical and real WTP estimates for each attribute. Each CF observation constitutes a unique pair of hypothetical to real WTP. For instance, suppose a study reported real and hypothetical WTP for two different goods, it would constitute two CFs. If it also reported median WTP values, it would generate two more CFs (four total). And if it also had a treatment and control group, this would again double the number of CFs. For example, Morrison and Brown (2009) used three samples for three treatments, cheap talk, dissonance minimization, and a control group, but provided a mean, median, and Turnbull lower bound estimate for each group,

yielding nine distinct observations. Loomis, et al. (1996) had three separate samples that generated four observations in our meta-analysis database (two within-sample and two between-sample comparisons) based on reported mean WTP and four more observations for the reported median WTP.

We take the natural log of CF (the same as LG), which consequently drops observations with non-positive CF.[8] MASW used the natural log of the real WTP[9] which can make interpretation somewhat counterintuitive. For example, their indicator for student samples decreases the log of the real WTP, which inherently increases the gap between hypothetical and real WTP. Our dependent variable is more naturally understood; if student respondents tend to overstate their WTP more than other respondents, then the predicted sign on a student dummy variable is positive. That is, all else equal, a student-only sample would lead to a higher CF. This dependent variable necessitates a stated and real WTP estimate per study as previously mentioned.

[8]A total of 3 observations were dropped because real WTP (the denominator of CF) equaled 0, making CF undefined; Alfnes et al. (ERAE, 2010, List and Shogren (2002) and Christie (2007)

[9]LG also used the absolute value of CF. We believe this may be unfavorable since it reduces differences in magnitude by treating a severe understatement of WTP as equivalent to a severe overstatement of WTP, since hypothetical WTP half the size of actual WTP, i.e. |ln(.5)|, would have the same value as an observation in which hypothetical WTP is twice as large as actual WTP, i.e. |ln(2)|.

Among the independent variables used in this study, several of them are adapted from previous meta-analyses. **Table 2-2** provides a description of each variable and their descriptive statistics. A more detailed description appears in Appendix C in the supplementary appendix online.

Previously considered variables that we follow nearly identically are as follows. We have indicators for whether the study valued a public good ("Public"),[10] if the sample was primarily composed of students ("Student"), and if the study used a split-sample/between-respondent or within-respondent design ("Between"). In the same spirit as LG, the year of publication ("Publication Year") is included to test if the magnitude of HB has changed over time. We differentiate studies that elicit willingness to accept values versus WTP (e.g. Bishop, et al. (1983)) ("WTA"),[11] as extensively reviewed by Horowitz and McConnell (2002)) and Tunçel and Hammitt (2014). Based on previous meta-analyses (List and Gallet, 2001, Murphy, et al., 2005), we expect Public, Students, Between, and WTA to increase CF. Because our awareness and techniques to deal with HB have increased, we expect CF to decrease with Publication Year.

[10] Future investigation could delineate quasi/pseudo-public goods from pure public goods. Ready, Champ and Lawton (2010) provide a good distinction of quasi-public goods as non-rival in that everyone can benefit from it once it is provided, but it is excludable in that the respondent will benefit from that unit of the good only if she makes the donation.

[11] Articles coded as WTA at least once are listed in Appendix D of the supplementary appendix online.

Beyond the mentioned variables, we have greatly expanded upon and created a number of new variables to capture differences in experimental protocols relative to LG and MASW. There may be potential differences in the survey delivery mode. With mail surveys as the baseline, we have three dummies for individual or group in-person lab settings, phone surveys, and field surveys, described under “Survey Mode.” In addition, we also control for whether the survey mode in the hypothetical valuation and real valuation are different. It is possible that some survey modes may not trigger HB as easily as other modes holding all other factors constant. As a result, depending on the different survey modes used in hypothetical and real valuations, the mismatched survey modes may attenuate or aggravate HB. Rather than attempt to characterize which hypothetical-real pairs are likely to mitigate or exacerbate HB, we include indicator variables for the most common survey mode mismatch, field-lab, and a second dummy to capture all remaining mismatches. Given our data, it is possible to capture any type of mismatches in survey modes but after some preliminary testing, other less popular mismatches are highly insignificant in the regression analysis suggesting we do not have enough observations to support identification. While many studies examine hypothetical WTP under multiple survey modes, as far as we have found, only Ethier, et al. (2000) studied the extent of HB in two survey modes (mail and phone), and found no difference.

Another major improvement are new variables (“HB Mitigation Technique”) to test for the relative effectiveness of various HB mitigation techniques. As in LBB, we distinguish CFs that used cheap talk and certainty follow-up, but also

add two more dummies for consequentiality and other HB mitigation techniques, in which the latter represents any other less popular mitigation technique. In most cases, these treatments were only studied in one or two studies, so sparse data makes modelling inappropriate.[12] The literature often separates mitigation techniques as either Ex Ante or Ex Post corrections. We explore this characterization with a separate model that includes two indicators to explicitly consider the efficacy of Ex Ante versus Ex Post HB mitigation techniques.[13] We expect that all HB mitigation variables should reduce CF.

The type of hypothetical elicitation technique used is included, described under "Elicitation Format" with five categories (most similar to LBB who had seven): all auction-type valuations (e.g. Vickrey, Random N^{th}, Smith, etc.), Dichotomous Choice, Referenda, Choice Experiment[14], and Open

[12] The literature review mentions many of these less explored (i.e. budget/substitute reminders) or newer mitigation techniques (i.e. honest priming, oath, etc.). Other methods not mentioned but still accounted for include real talk (Alfnes Yue and Jensen, 2010), payment anonymity (Alpizar, Carlsson and Johansson-Stenman 2008) and payment immediacy (Veisten & Navrud, 2006), among others.

[13] Appendix E, available in the supplementary appendix online, details which articles have HB mitigation techniques included in "Other HB Mitigation Techniques" as well as how all HB mitigation techniques were assigned to either the Ex Ante or Ex Post HB mitigation in the subsequent model variant.

[14] These variables appeared in LBB's meta-analysis on the presence or absence of HB, but are unstudied in a meta-analysis on the magnitude of HB.

Ended/Payment card elicitation, yielding four dummy variables.[15] With regard to which elicitation mechanism may generate the most or least HB, the NOAA Blue Ribbon Panel believed in the credibility of Dichotomous Choice compared to an Open-Ended approach, but generally the literature is mixed (see Table 1 of Champ and Bishop (2006) for a review). As a result, we hope our study offers additional evidence to this discussion.

Many studies use the non-hypothetical treatment as the basis for real WTP, but even real WTP can be a function of the elicitation mechanism (Champ and Bishop, 2006, Gracia, et al., 2011, Lusk and Schroeder, 2006), justifying MASW's decision to exclude studies that use different elicitation mechanisms in the real treatment. Instead, similar to the mismatching survey mode dummies, we include observations that have mismatching elicitation mechanisms using two indicators: one for the most common mismatch, hypothetical Open-ended/Payment Card and a real Auction, and a second to indicate remaining mismatching elicitation observations. For a similar reason as in the survey mode mismatches, the effect from other types of mismatched elicitation methods on CFs cannot be identified.

Finally, we include indicator variables for studies using induced-values (e.g. Taylor, et al. (2001) and Mozumder and Berrens (2007)) ("Induced

[15] LG had 8 categories (OE, 1st price sealed bid, Vickrey 2nd price auction, provision point, Smith auction, random price auction, BDM, and DC) while MASW only had 1 distinction ("Choice" which includes dichotomous or polychotomous choice, referendum, payment card, and choice experiments versus auctions).

Value")[14,16] and if they are peer-reviewed ("Peer-Reviewed"). Whereas some studies provide no money in the real treatment (Michaud, et al., 2013), others provide some sort of participation incentive. To test for "Found Money" effects (Cummings and Taylor, 1999), we use a dummy for presence or absence of an endowment ("Endowment") given to respondents, whether hypothetical or real.[17] To test for a CF's potential sensitivity to the type of measures used, we add two controls for Median WTP and Turnbull lower bound WTP ("Median" and "Turnbull"), and the omitted reference group is comprised of both mean WTP and WTP estimated directly from the model. In terms of expectations, because values are assigned, Induced Value studies should have lower CF (Mitani and Flores, 2014). While it has been shown that endowments affect bids and WTP, we have no expectation of endowment in CF because no study has shown the endowment effect in a hypothetical versus a real setting. By including non-peer-reviewed manuscripts, we avoid Publication Bias (Stanley, 2008), though there is no reason to expect it to affect CF. Similarly, there is no reason to suspect that median or Turnbull measures should adversely affect CF.

A number of other potential characteristics exist, such as distinguishing charitable donations from purchases (e.g. Brown, et al. (1996) and Macmillan, et

[16] CF in induced value observations is calculated based on the observed hypothetical compared to the observed real value, rather than the induced value itself, i.e. the actual value assigned to the respondent.

[17] At this time, we do not distinguish studies that provided starting funds using techniques to mitigate "found money" effects.

al. (1999)), or delineating target product categories examined in a study such as food, the environment and health. All of the articles included in our meta-analysis sample admittedly ignore such details and information. The constant in a meta-analysis regression model encapsulates the reference categories for each independent variable.

2.5 Model Specification

To understand HB, equation (1), a "fixed-effect-size" meta-regression model according to Nelson and Kennedy (2009) is used such that the natural log of Calibration Factor is a function of the variables defined earlier:

$$\boldsymbol{ln.CF} = \boldsymbol{ln}\left(\frac{\boldsymbol{Hypothetical\ WTP}}{\boldsymbol{Actual\ WTP}}\right) = f(\textit{Public, Student, Between, Publication Year, WTA, Survey Mode, Field-Lab Mismatch, Other Survey Mismatch, HB Mitigation Technique (Ex Ante/Ex Post mitigation), Elicitation Format, OEPC- Auction Mismatch, Other Elicitation Mismatch, Induced Value, Peer Reviewed, Endowment, Median WTP, Turnbull}) \quad (1)$$

A number of specifications and robustness checks were used to ensure model validity. In addition to estimating the natural log of CF, the model was also run without the logarithm transformation, as well as the absolute value of the natural log of CF, as in LG. Transforming the dependent variable did not substantively affect the results, including the sign or significance of the coefficients. As will be seen later, the untransformed data is also highly skewed to the right such that a transformation makes data more normally distributed. A Box-Cox test provided statistical evidence favoring a log transformation.

In addition to the full models, we estimate trimmed models that eliminate 5% of the observations, those with a CF outside the 2.5 and 97.5 percentiles. MASW use a similar method, serving as a check on the sensitivity of the results to the most extreme observations.

As stated before, some studies (for example Loomis, et al. (1996) and Morrison and Brown (2009)) report multiple measures of WTP, such as mean and median WTP, per good. In these cases, each CF is added as a separate observation into the meta-analysis dataset, corresponding to each measure. Accordingly, the multiple CFs are reweighted.[18] For example, if a paper provides two CFs based on mean and median WTP, both CFs enter the sample as two separate observations, but each observation is weighted by 0.5. Reweighting observations is also especially important to studies employing a choice experiment. For instance, if the same sample of respondents produces five CFs for five attributes,[19] each CF is weighted by 0.2. Lastly, we use cluster-robust standard errors to allow for correlation across observations based on the same study. Some authors (e.g. Champ (Champ, et al., 2009, Ready, et al., 2010), Blomquist (Blomquist, et al., 2009, Blumenschein, et al., 1998), among others)

[18] Different goods are treated as separate samples. By default, all studies with 1 row are not reweighted.

[19] WTP is typically defined as -1*(attribute coefficient/price coefficient). For some results, it is only necessary to include one attribute since the CF will be equal across attributes. In these cases, we only use one attribute since adding multiple CF's and subsequent weights is redundant.

are responsible for multiple studies such that clustering on certain authors is also reasonable,[20] but this also had no meaningful effect on the results.

2.6 Results

Of over 280 articles considered, 132 met the necessary requisites for inclusion, generating a total of 908 observations[21] for the meta-analysis.[22] The number of corresponding observations per study varied depending on the amount of useable WTP information available. The mean (median) number of observations in the meta-analysis data generated from one study was 6.82 (4), while the minimum and maximum for any study were 1 (several studies) and 71 (Alfnes, et al., 2010)), respectively. Once weights are included, the effective sample size is approximately 336.74. The trimmed sample removed observations

[20] A list of clusters on authors is provided in Appendix G in the supplementary appendix online. Another potential clustering method is based on the same dataset rather than just publication. For example, Ethier et al. (2000), Poe et al. (2002), and Cameron et al. (2002) appear to all be based on the same data. This different clustering method was not attempted in this analysis, but can be readily examined in a future study.

[21] This includes 21 observations that have CF≤0, which are inherently dropped in ln(CF) models.

[22] An explanation of each study's exclusion appears in Appendix A and a complete list of the excluded articles appears in Appendix B, both available in the supplementary appendix online.

with the 23 smallest and 23 largest CFs, approximately 5% of the sample, for 862 observations and effective sample size of 324.42.

HB is quite pervasive throughout the sample as demonstrated in **Figure 2-1**. The mean (median) CF of the dataset is 2.29 (1.39). In the full dataset, the minimum CF is -37.10 and the maximum CF is 48.39, which gives a much greater range of the data relative to MASW, who reported a minimum (maximum) CF of 0.76 (25.08) as well as LG, whose minimum (maximum) CF was 0.5 (28.2). In the trimmed sample the CF is more moderate, with a mean of 1.94 (same median) and a minimum and maximum CF of 0.08 and 13.00, respectively. CFs based on observations without any form of HB mitigation are also displayed. It shows that roughly one of every four observations has hypothetical WTP between 81% and 120% of real WTP, and roughly another quarter of observations have hypothetical WTP between 121% and 160% of real WTP without additional HB mitigation methods implemented.

Model results based on the natural log of CF appear in **Table 2-3**. It starts with differences in implementation across studies such as public good, student respondents, etc., and continues with elicitation mechanism, survey mode, and HB mitigation strategies. The table presents four models, a model specifying the specific HB mitigation techniques, labeled “HB”, the Ex Ante/Ex Post variant, and both models in the full and trimmed samples.[23] With R^2 at a range of 0.21 to

[23] The results of the 4 model variants using a linear specification of CF are available in Appendix I. The linear specification allows for the incorporation of observations with

0.28, the goodness of fit in our models is less than LG (R^2 range of 0.4 to 0.5) and much less than MASW's explanatory power of 0.83 to 0.87.[24] In general, most results were robust when using the trimmed sample. As expected, eliminating the most extreme observations reduces the effect of each independent variable. We refer to variable effects on CF and HB interchangeably.

In discussing the findings, we first focus on variables that have been looked at in previous meta-analyses and proceed to variables that have not be investigated in the past. As a whole, a number of study characteristics are significant, indicating that decisions in the study implementation process can indeed affect the presence and magnitude of HB. Particular variable results relative to previous conclusions are mixed. For significant variables, we provide the marginal effect of the variable. For example, as in LG and MASW, based on our results, public goods have higher CFs in all models, generating 84.7% and 64.9% higher CFs in the full and trimmed sample, respectively (for the "HB"

CF≤0. The trimmed linear specification drops these same observations and is identical to the trimmed ln(CF) model. Regardless, results between the linear and logged specifications are nearly identical in the full and trimmed samples.

[24] We suspect that the reason for the higher fit in MASW is their inclusion of [the natural log of] hypothetical WTP as well as a squared term as explanatory variables of ln(actual WTP). For example, the adjusted R^2 in their baseline model (Model 1a, pg. 319) was 0.83, with ln(Hypothetical WTP) and its square alone. We used the information of hypothetical WTP to construct the CF in our analysis.

specifications in **Table 2-3**). We find no evidence that student respondents affect the magnitude of HB, unlike MASW and LBB, which supports researchers' continued use of student respondents. This still does not imply using student samples to infer to broader populations, but it appears using student samples may be well suited for tests of economic theory or methodology without the concern of involving more HB than using a general public sample. Between-respondent designs do not affect the CF, which coincides with previous findings except for MASW. As LG point out, this means that a between-respondent design does not inherently bias results, but the ideal is still a within-design in order to reduce the number of potential confounding factors across treatments. Publication year is not significant, similar to what is briefly mentioned in LG (Footnote 9, pg. 252). The time trend of CF in our study has a positive sign, as was found in LG, whereas LBB found a negative sign, though in all three cases still not significant.

The indicator for CFs based on WTA was sensitive to the use of a full dataset, in which it was not significant, versus a trimmed dataset, in which WTA is associated with a *lower* rate of HB. This unusual result is driven by the fact that among the 23 observations (97.5 percentile) dropped, 12 were WTA studies, all stemming from Brookshire and Coursey (1987), which also represents 23.5% of all 51 observations coded as WTA. Only one WTA observation was dropped from the smallest 2.5% of observations. On the other hand, 26 of the 51 WTA observations had CFs between .5 and 1.5.

Inspecting the effect of survey mode shows that relative to a traditional mail survey, lab, phone-based, field surveys/experiments, and online surveys had no effect on CF. Again, these findings show that the survey mode is not a major contributor for HB. Researchers may have more confidence choosing the mode of survey that best fits the needs and circumstances of their particular research.

In reference to various HB mitigation techniques implemented, cheap talk, certainty follow-up, and consequentiality were negative and statistically significant, providing evidence of the value of such strategies to reduce HB. Certainty follow-up questions had similar effects, with certainty follow-up reducing CF by 136% and 99% in the full and trimmed samples, and consequentiality reducing CF by about 137% and 95% reduction for the respective samples. With respect to certainty follow-up, we also conducted a related investigation distinguishing observations into those that used qualitative certainty follow-up (e.g. “Very Likely”) and those that used quantitative certainty follow-up (e.g. 1-10 scale). In every model specification we examined, this differentiation does not show any statistical significance. Cheap Talk reduced CF by 70% in the full sample and by about 41% in the trimmed model. This seems to match the mixed usefulness of cheap talk found throughout the literature. The indicator for other HB mitigation techniques was not significant in the full or trimmed models.

For elicitation method, we find that Choice Experiments and Referendums generate significantly lower CF compared to the reference category, dichotomous choice methods. CF for choice experiments is approximately 60%

lower in choice experiments and 95% lower in referendums[25] with minimal differences in the full and trimmed samples. Given its important historical context, it is interesting to see that Open-Ended/Payment Card approaches are not significantly different than Dichotomous Choice. The model results coincide with MASW, who found that "choice" (composed of DC, referendum, payment card and CE) mechanisms reduced HB compared to various auction types, but opposite of LG who largely found elicitation mechanisms were not factors. Lastly, mismatching elicitation mechanisms do not appear to affect CF.

Turning our attention to the results of unexplored variables in previous meta-analyses, unsurprisingly, induced value experiments are consistently significant and reduce CF, giving one of the biggest effects on decreasing the magnitude of HB across all four models. The control for peer-reviewed publications does not show evidence of a difference in the magnitude of HB. Endowment was also not statistically significant in any model. Much like the controversy over the use of students, this evidence may reduce concern for potential bias introduced by use of participation fees. The variables associated with mismatching survey modes and mismatching elicitation mechanisms in the hypothetical and real valuations provide useful information. The full sample shows that a hypothetical field-real lab approach generates CF much larger than a matching design. This mismatch is not significant in the trimmed sample, driven by the fact that 13 of 23 removed observations were Field-Lab mismatches. On

[25] This result is contrary to Polome's (2006) comparison of the referendum and dichotomous choice elicitation methods.

the other hand, whether the elicitation mechanisms match between the hypothetical and real valuation elicitation does not appear to introduce any additional HB into the results. Finally, observations based on median WTP significantly increased CF by about 80%, but only in the full sample. Turnbull WTP observations did not significantly affect CF.[26] This shows that observations based on the non-parametric Turnbull approach are not systematically different, yet can still add to the size of the dataset. Overall, these results do suggest differences in the prevalence of HB based on the experimental protocols and other study-specific characteristics, but such conclusions should be considered cautiously since some results were sensitive to the trimmed versus full datasets.

For the model variant that recharacterized all HB mitigation techniques as either Ex Ante or Ex Post approach, the results are similar to the individual HB counterpart. Ex Post, which includes certainty follow-up and calibration, significantly reduces CF by 100% and 73% in the full and trimmed samples. On the other hand, Ex Ante approaches are significant but have about half the effect, and only marginal significance in the trimmed model. Since about two-thirds of the observations coded as “other HB mitigation technique”, which was not statistically significant, were recoded as Ex Ante methods, this may explain part of the divergence Ex Ante’s benefit in the trimmed sample.

[26] Per a reviewer recommendation, we further divided Turnbull-based WTP observations into those based on a single-price and those based on multiple offers. This extension was not statistically significant under multiple model specifications and is therefore not reflected in the final model.

Among the elicitation methods, survey mode, and HB mitigation techniques, we applied a series of post-estimation Wald tests of the equality for parameter estimates. These tests were implemented across all four model variants. The results of the elicitation mechanism and HB mitigation techniques appear in **Table 2-4** and **Table 2-5**, respectively. These results provide statistical evidence that correspond to the regression results. For instance, **Table 2-3** shows that studies with auctions have higher CF relative to dichotomous choice studies, whereas all of the other elicitation mechanisms have significantly lower CF compares to dichotomous choice studies. Accordingly, **Table 2-4** shows that each of these other elicitation mechanisms also have significantly lower CF compared to auction methods. Lastly, choice experiments, referendum, and open-ended/payment card approaches all have smaller CFs, they are statistically equivalent, consistent with the findings of Cameron, et al. (2002), with a possible exception between referenda and open-ended/payment card.

Comparing the HB mitigation techniques of **Table 2-5**, it is important to remember that all techniques had a negative sign, indicating they reduced HB compared to no mitigation implemented, so these results compare which techniques are more or less effective at reducing HB. There is statistically significant evidence that certainty follow-up and consequentiality reduce HB more than cheap talk, but they are not significantly different from each other. This gives credence to the promise of consequentiality to both explain and reduce HB. Lastly, the coefficient estimate of Other HB techniques was significantly greater than for the three main HB mitigation techniques.

We also tested equality of Ex Ante vs Ex Post CF mitigation as well as the various survey modes (not presented). The coefficient of Ex Post techniques to reduce CF was marginally significantly better than Ex Ante techniques in both the full (p-value=.080) and trimmed (p-value=0.056) models. In all model variants, there was no statistical significance to suggest a difference in individual survey modes.

2.7 Discussion

The prominence of stated preference methods continues to grow in multiple fields of economics and other disciplines. While tools to alleviate HB have been developed, there is no consensus in the literature on either what theories best describe why HB persists or which tool may function most efficiently on average to reduce HB. We provide an updated and augmented meta-analysis, both in its comprehensiveness of studies and variables compiled, to provide new evidence in the discussion of HB. In this process, we also investigate and verify previous meta-analyses' findings on the importance of various study characteristics and consider new factors as potential determinants of HB. Coincidentally, even with considerably more studies included, the average CF in the trimmed sample was 1.94, closely corresponding to previous analyses of HB.

Overall, some results are sensitive to using a full or trimmed sample, specifically for hypothetical Field-real Lab survey mode mismatch, CF based on median, and WTA. We find that the significance and effect of public goods and induced value experiments as well as the lack of effect for students, between-

respondent designs, endowments, and mismatching elicitation modes are consistent between the full and the trimmed samples. Lastly, HB mitigation techniques, characterized individually (i.e. cheap talk, certainty follow-up and consequentiality) or as Ex Ante and Ex Post, do work in the reduction of HB. Overall, these findings can inform researchers the degree to which their results of stated WTP should be adjusted upward or downward to come closer to real WTP. Alternatively, for those in the formulation stage of a stated preference valuation, these results may suggest how much they should be concerned with HB and adjusting mitigation techniques.

Our results are not without limitation. While considerable effort was made to compile the most accurate meta-analysis data possible, other perspectives and definitions of variables may change the results and implications. In addition, the dataset is a product of tedious, but fallible hand coding. And even with a richer dataset, the ability to explain divergence between stated and real WTP remained low throughout all of the models. Regardless, some of the results that are least susceptible to these misgivings are also the most striking. Some of the most promising results are the absence of student sample effects, the consistent usefulness (and magnitude) of HB mitigation techniques, and the importance of elicitation mechanisms used.

While we aim to corroborate and update previous results, we see the same opportunity in our work. Additional investigation would benefit from more variable refinement and further separation of other potential determinants. Moving forward, there are multiple extensions and questions to investigate using

this much richer dataset. For instance, distinguishing quasi-public goods, which maintain some degree of rivalry from public goods, categorizing goods as either for health, environmental or food, or indicating the use of two or more HB mitigation techniques in one study, may all be proven to play some part in dictating the amount of HB. In addition, the potential to more finely characterize payments such as ‘earned money’ or outright ‘participation fees’ across studies can shed more light on the impact of endowment effects. Another opportunity for possible extension is to include a measure of study quality (Loomis, 2011). As stated preference approaches continue to grow in use and in different fields of research, we can expect more studies to benefit from a meta-analysis like ours and try to understand the difference between hypothetical and real values.

Table 2-1 Select Results from Previous Meta-Analyses on Hypothetical Bias

Study	List & Gallet (LG) (2001)	Little & Berrens (LB) (2004)	Murphy, Allen, Stevens & Weatherhead (MASW) (2005)	Little, Broadbent & Berrens (LBB) (2012)
Dependent Variable	ln(Hyp. WTP/Real WTP)	1= HB present, else 0	Ln(Act)	1= HB present, else 0
# of Studies (Observations)	29 (58)	53 (85)[a]	28 (77)	96 (220)
Private Good	SS, Less HB	Not SS, Less HB	SS, Less HB	--[b]
Student Sample	--	--	SS, More HB	SS, More HB
Within Respondent	Not SS, Less HB	Not SS, Less HB	SS, Less HB	Not SS, More HB
Willingness to Accept	--	Not SS, More HB	--	Not SS, More HB
Lab Setting	Not SS, Less HB	Not SS, More HB	--	Not SS, Less HB
HB Mitigation Techniques	--	--	SS, Less HB	SS, Less HB[c]
Choice Experiment	--	--	--	Not SS, Less HB
Induced Value	--	--	--	SS, Less HB

Note: SS: Statistically Significant, Not SS: Not Statistically Significant. Results are considered significant if the variable was marginally significant in 50% or more of the appropriate models. "--" indicates the variable was not included in the meta-analysis' model.

[a] Based on information reported in LBB 2012.

[b] LBB model distinguish the differences of studies by modeling interactions of public and private goods with other study characteristics; no variable specifically models public and private good differences.

[c] LBB modeled cheap talk and certainty follow-up as separate variables, but both were significant and had the same (negative) sign.

Table 2-2 Variable Description and Sample Proportion[1] (n=908)

Variable	Prop.[2]	Variable Description [Reference Category]
Public Good	.385 .615	1 if the good or service under consideration is a public good, else 0 [private]
Student	.358 .642	1 if respondents are primarily made up of students, else 0 [Non-students such as the general population, shoppers, etc.]
Between-Respondent	.804 .196	1 if respondents are different in the hypothetical and actual treatments, else 0 [same respondents in the hypothetical and actual treatments]
Publication Year	NA	Discrete continuous variable indicating the year the paper was published. Min= 1 (1974), Max= 43 (2014)
Willingness to Accept	.056 .944	1 if study uses Willingness to Accept, else 0 [Willingness to Pay]
Hypothetical Survey Mode		
Lab	.664	1 if individual or group in-person (i.e. lab setting),
Phone	.027	1 if phone survey,
Field	.202	1 if field survey/experiment,
Online	.010	1 if online survey,
	.097	else 0 [mail survey]
Mismatching Survey		
Field-Lab Mismatch	.038	1 if the hypothetical-real mismatch is field-lab
Other Survey	.042	1 if any other pair of hypothetical-real mismatching survey modes
Mismatch	.920	else 0 [hypothetical and actual survey modes are the same]
A. HB Mitigation Technique[2]		
Cheap Talk	.117	1 if cheap talk,
Certainty Follow-up	.046	1 if certainty follow-up,
Consequentiality	.010	1 if consequentiality,
Other HB Mitigators	.129	1 if any other HB mitigation technique,
	.708	else 0 [no mitigation technique used]

Table 2-2 Continued Variable Description and Sample Proportion[1] (n=908)

B. Type of HB Mitigation Technique		
Ex Ante	.198	1 if Ex Ante,
Ex Post	.103	1 if Ex Post,
	.708	else 0 [no mitigation technique used]
Hypothetical Elicitation Format		
Auction	.254	1 if auction (i.e. Vickrey, N^{th} Price, BDM, Smith, etc.)
Choice Experiment	.306	1 if choice experiment,
Referendum	.091	1 if referendum[3],
Open-End/PCard	.191	1 if open ended or payment card,
	.158	else 0 [dichotomous choice]
Mismatching Elicitation		
OEPC-Auction	.070	1 if the hypothetical-real mismatch is OEPC-Auction
Mismatch	.036	1 if any other hypothetical-real mismatching elicitation mechanisms
Other Elic Mismatch	.893	else 0, [hypothetical and actual elicitation mechanisms are the same]
Induced Value	.057	1 if Induced Value (when respondents are assigned their values),
	.943	else 0 [Respondent's homegrown WTP]
Peer Reviewed	.965	1 if peer-reviewed publication
	.035	else 0 [not peer-reviewed]
Endowment	.690	1 if the respondent receives a financial participation incentive,
	.310	else 0 [No money given]
WTP Type		
Median	.149	1 if median WTP, else 0
Turnbull	.141	1 if Turnbull lower bound WTP,
	.710	else 0 [mean/model estimate WTP]

[1]A more descriptive definition and example studies are provided in Appendix C of the supplementary appendix online.

[2]Proportions may not sum to 1 since some studies use multiple mitigation techniques in one treatment.

[3]Distinct from dichotomous choice since it relies on a group vote.

Table 2-3 Model Results of Study Characteristics' Effect on ln(Calibration Factor)

	Full Dataset				Trimmed Dataset			
	HB	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**	**HB**	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**
Public Good	0.35*** (0.13)	84.7%	0.349** (0.133)	84.1%	0.306** (0.129)	64.9%	0.304** (0.132)	64.3%
Student	-0.031 (0.157)		-0.026 (0.157)		-0.072 (0.149)		-0.067 (0.149)	
Between-Respondent	-0.02 (0.111)		-0.046 (0.114)		-0.029 (0.101)		-0.054 (0.103)	
Publication Year	0.004 (0.007)		0.004 (0.008)		0.003 (0.007)		0.003 (0.007)	
Willingness to Accept	-0.144 (0.34)		-0.156 (0.342)		-0.358** (0.147)	-58.9%	-0.375** (0.154)	-61.2%
Survey Mode								
Lab	0.076 (0.244)		0.093 (0.24)		0.086 (0.235)		0.102 (0.231)	
Phone	0.143 (0.242)		0.119 (0.23)		0.122 (0.225)		0.104 (0.214)	
Field	0.069 (0.189)		0.068 (0.198)		0.031 (0.182)		0.035 (0.19)	
Online	0.13 (0.256)		0.103 (0.262)		0.079 (0.246)		0.048 (0.25)	
Field-Lab Mismatch	0.729* (0.395)	227.9%	0.787** (0.394)	253.8%	-0.042 (0.329)		0.006 (0.327)	
Other Survey Mismatch	0.196 (0.312)		0.194 (0.306)		0.18 (0.304)		0.174 (0.298)	

Table 2-3 Continued Model Results of Study Characteristics' Effect on ln(Calibration Factor)

	Full Dataset				Trimmed Dataset			
	HB	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**	**HB**	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**
HB Mitigation Technique								
Cheap Talk	-0.297*** (0.087)	-69.5%			-0.231*** (0.086)	-40.5%		
Certainty Follow-up	-0.698*** (0.125)	-135.7%			-0.696*** (0.12)	-98.5%		
Consequentiality	-0.709*** (0.133)	-137.3%			-0.66*** (0.126)	-95.0%		
Other HB Mitigators	-0.065 (0.114)				-0.064 (0.123)			
Ex Ante			-0.19** (0.082)	-46.8%			-0.157* (0.084)	-28.6%
Ex Post			-0.463*** (0.151)	-100.2%			-0.46*** (0.153)	-72.5%

Table 2-3 Continued Model Results of Study Characteristics' Effect on ln(Calibration Factor)

	Full Dataset				Trimmed Dataset			
	HB	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**	**HB**	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**
Elicitation Format								
Auction	0.213 (0.166)		0.215 (0.178)		0.107 (0.152)		0.108 (0.164)	
Choice Experiment	-0.353** (0.163)	-64.9%	-0.352** (0.169)	-64.8%	-0.33** (0.149)	-61.4%	-0.324** (0.156)	-60.3%
Referendum	-0.637*** (0.184)	-102.7%	-0.63*** (0.209)	-102.0%	-0.579*** (0.176)	-95.9%	-0.57*** (0.199)	-94.7%
Open-End/ PCard	-0.226 (0.155)		-0.158 (0.156)		-0.235 (0.148)		-0.172 (0.148)	
OEPC-Auction Mismatch	0.280 (0.228)		0.209 (0.238)		0.182 (0.222)		0.119 (0.23)	
Other Elic. Mismatch	0.003 (0.255)		0.026 (0.255)		0.013 (0.25)		0.034 (0.25)	

Table 2-3 Continued Model Results of Study Characteristics' Effect on ln(Calibration Factor)

	Full Dataset				Trimmed Dataset			
	HB	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**	**HB**	**Marginal Effect**	**Ex Ante/ Ex Post**	**Marginal Effect**
Induced Value	-0.800*** (0.230)	-141.4%	-0.765*** (0.231)	-137.3%	-0.693*** (0.192)	-98.9%	-0.661*** (0.193)	-95.7%
Peer Reviewed	0.148 (0.203)		0.122 (0.200)		0.098 (0.198)		0.07 (0.192)	
Endowment	0.075 (0.142)		0.014 (0.143)		0.032 (0.129)		-0.028 (0.13)	
Median WTP	0.291* (0.167)	78.8%	0.295* (0.173)	80.0%	0.186 (0.133)		0.184 (0.14)	
Turnbull WTP	0.158 (0.155)		0.097 (0.157)		0.123 (0.147)		0.064 (0.149)	
Constant	-8.383 (14.811)		-8.12 (15.201)					
R^2	0.28		.26		0.23		.21	
N	**887**		**887**		**862**		**862**	
(Weighted N)	**(332.96)**		**(332.96)**		**(324.42)**		**(324.42)**	

Cluster (per study) Robust Standard Errors reported in parentheses. Asterisks ***, **, and * indicate a p-value <.01, <.05, and <.1, respectively.

Note: The dependent variable, Calibration Factor, is the natural log of the ratio Hypothetical WTP divided by Actual WTP.

Table 2-4 P-values of Wald Tests of Elicitation Methods

	Auction	**Choice Experiment**	**Referendum**	**Open-End/ Payment Card**
Auction	--			
Choice Experiment	Less than 0.001<	--		
Referendum	Less than 0.001<	0.138	--	
Open-End/ Payment Card	0.023<	0.542	0.053>	--

Note: < (>) indicates that the coefficient of the row variable is significantly smaller (larger) than the coefficient of the corresponding column variable. Test results based on full sample, but were equivalent in the trimmed sample with the exception of referendum versus OEPC, which was not significantly different.

Table 2-5 P-values of Wald Tests of HB Mitigation Methods

	Cheap Talk	Certainty Follow-up	Consequentiality	Other HB Techniques
Cheap Talk	--			
Certainty Follow-up	0.006<	--		
Consequentiality	0.008<	0.942	--	
Other HB Techniques	0.076>	Less than 0.001>	Less than 0.001>	--

Note: < (>) indicates that the coefficient of the row variable is significantly smaller (larger) than the coefficient of the corresponding column variable.
Test results based on full sample, but were equivalent in the trimmed sample, with the exception of Other HB, which was not significantly different from Cheap Talk.

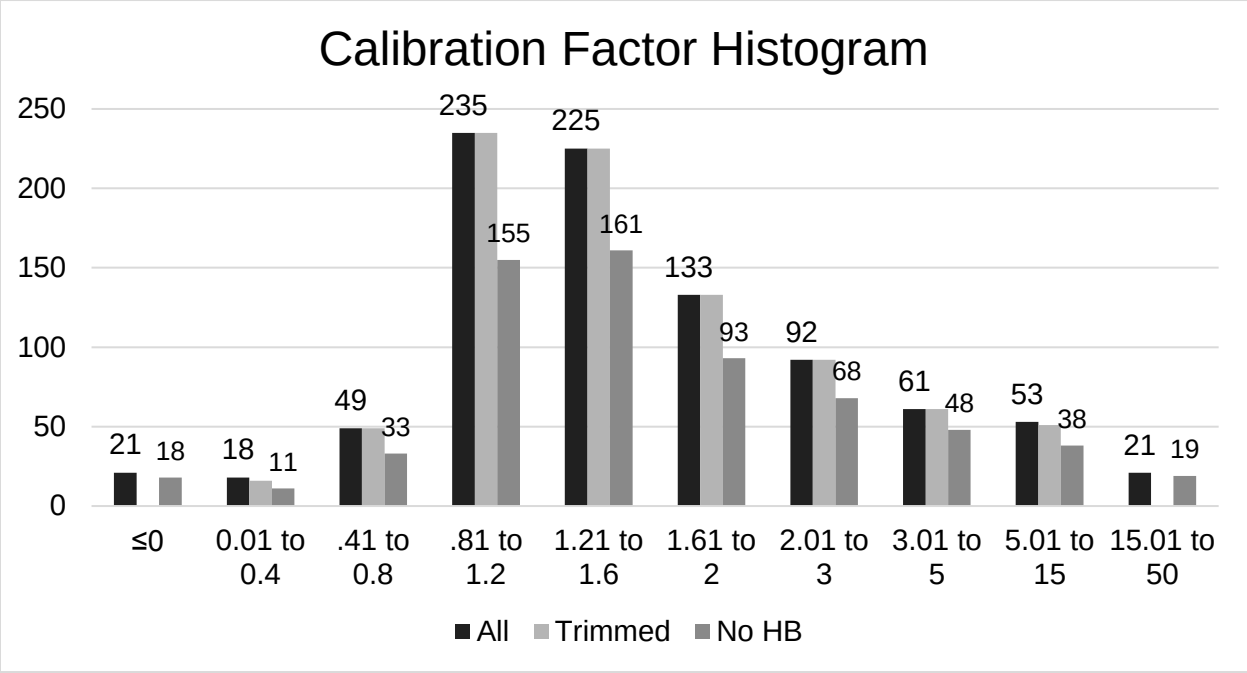

Figure 2-1 Histogram of sample's calibration factors (n=908)

Chapter 3 THE PRESENCE OF HYPOTHETICAL BIAS WITHIN SPATIAL DECAY AND CHARISMATIC SPECIES: AN APPLICATION OF MONARCH AND VICEROY BUTTERFLIES

3.1 Abstract

Researchers have regularly used stated preference methods to study species valuation and more recently to investigate spatial heterogeneity/distance decay in welfare estimates. Yet, Hypothetical Bias (HB) is an ongoing concern for stated preference methods. In this analysis, we investigate the presence of HB within distance decay in a choice experiment of monarch and viceroy butterflies. Further, monarchs and viceroys are similar except that the former is well known and at-risk, while the latter is unfamiliar but common. This comparison enables the identification of a specific form of value associated with rare species, which we term a charisma effect, and the extent of HB due to the charisma effect. Results show that there is HB and distance decay in value for both butterfly species, but HB in distance decay is only found for monarchs and not for viceroys. We find that a charisma effect for monarchs exists in the hypothetical valuation scenarios, but disappears when the valuation involves real payment. Using our results to modify previous investigations of rare species generates lower, more believable welfare estimates.

3.2 Introduction

Within environmental economics, observing non-use values is difficult either through markets directly or through revealed preference mechanisms. This has necessitated stated preference methods such as Contingent Valuation or Choice Experiments (CE). However, stated preference methods regularly generate welfare estimates, such as Willingness to Pay (WTP) measures, greater than what one would observe in a non-hypothetical situation, with the difference commonly known as Hypothetical Bias (HB).

A separate vein of research within stated preference approaches that has recently garnered attention is the recognition of spatial heterogeneity of welfare estimates such as distance decay or hotspots. Distance decay is the circumstance in which the value of a species or environmental site decreases as the person's physical distance from species or environmental site increases, all else held constant. Hotspots and patchiness refers to local spatial patterns beyond continuous homogeneity or continuous decay (Johnston and Ramachandran, 2014). A number of studies employing stated preference methods document the presence of spatial decay or patchiness and its potentially large impact on aggregate welfare estimates.

As far as we know, the few studies that consider distance decay focus on iconic species that are available only within a relatively small range of geographical location instead of rare but largely distributed species. Furthermore, no studies examining the existence of geographic impacts on WTP have included an elicitation mechanism involving actual payment, and therefore there has been

no assessment of the extent of HB with respect to spatial decay. We investigate the presence of HB in a study of spatial decay using an application of butterflies.

This application of butterflies leads to the second primary contribution of this study. In the United States, monarchs (*Danaus plexippus*) are one of the most well-known butterflies, easily recognizable due to their vibrant orange color pattern. Recently, the monarch population has plummeted to a fraction of its former size, so much so that its restoration was included as one of three primary goals in the "National Strategy to Promote the Health of Honey Bees and Other Pollinators" (Pollinator Health Task Force, 2015). Further, monarchs are currently under status review for inclusion on the endangered species list (US Fish and Wildlife Service, 2016).

Monarchs could be considered a charismatic species, one that is well known and recognizable by the public and used for broader conservation initiatives. However, Brown and Shogren (1998) suggested that such well-known species generate "suspiciously high" values, such that "less than 2% of all threatened and endangered species represented 1% of the 1995 US GNP," evidence of HB. Our study design allows us to investigate a potential increase in HB due to charisma, a first for HB on studies of threatened, endangered, or rare (TER) species. We achieve this through comparing the values of monarchs to the viceroy butterfly (*Limenitis archippus*), which is nearly identical in its shape and appearance.

3.3 Background

3.3.1 Distance Decay

Studies on spatial decay were borne out of the need to generate more accurate welfare estimates of resources and amenities by including all relevant populations, especially those outside of the immediate vicinity of a resource. A resource's total economic value is understated if non-zero values of people from more distance locations are excluded from the analysis, but may be overstated if the value is assumed to be equal to those closest to the resource. Sutherland and Walsh (1985) were among the earliest to document this negative relationship between value and distance, and studies continue to consider distance decay either in use or non-use values (del Saz Salazar and Menéndez, 2007, León, et al., 2016, Schaafsma, et al., 2012).

Rolfe and Windle (2012) outline four principle reasons for declining values over distance: 1) use value declines as people live further away, 2) more or different substitutes become available as distance increases, 3) less ownership/responsibility for more distant environmental assets in different locations, and 4) lower awareness and knowledge of more distant environmental assets (Hanley, et al., 2003, Pate and Loomis, 1997, Sutherland and Walsh, 1985).

Recently, efforts have shifted from spatial uniformity or simple linear distance decay to whether there is spatial correlation in local areas that affect WTP. Johnston and Ramachandran (2014) found that most attributes did not exhibit global distance decay, but still found significant heterogeneity in WTP at

the local level, termed patchiness. Campbell, et al. (2008) and Meyerhoff (2013) both find evidence of local spatial clustering in WTP. Recently, Johnston, et al. (2015) showed the importance of spatial clustering as it relates to the spatial scale under consideration (e.g. gathering responses within 50km versus 500km of a particular site), and that cold and hot spot WTP patterns can change with the spatial scale.

Yet, with the mounting evidence of spatial heterogeneity in WTP, so far, these values are obtained through surveys involving hypothetical valuation questions. In other words, respondents to these surveys do not have to actually pay what they indicated in the survey—a situation that could generate HB. It seems imperative to test the extent of Hypothetical Bias for these same measures.

HB in valuation is the difference between a welfare estimate, usually WTP, that stems from a hypothetical elicitation in which the respondent's decision has no real payment consequence, and a real elicitation, in which payment is binding. Multiple meta-analyses have noted the consistent upward bias and its relevance across a variety of fields and types of goods and services (List and Gallet, 2001, Murphy, et al., 2005). To study this issue, we implement a real and hypothetical Choice Experiment on the willingness to support butterfly conservation in multiple locations involving different distances to the site of conservation. The application on butterflies also yields our second contribution, the charisma effect, as outlined below.

3.3.2 Butterflies and Charisma

Monarchs are one of the most well-known butterflies in the United States, easily recognizable from its vibrant orange color pattern and its annual migration across North America. For a number of reasons, the monarch population has plummeted to a fraction of its observed size since tracking began in the mid-90's (Brower, et al., 2012, Jepsen, et al., 2015). The monarch butterfly was initially placed under status review for inclusion on the endangered species list (Kaufman, 2014) in 2014, with a final decision due in 2019 (US Fish and Wildlife Service, 2016). The Obama administration acknowledged this collapse in its release of the National Strategy to Promote the Health of Honey Bees and Other Pollinators (Pollinator Health Task Force, 2015). One of its primary goals is to "increase the Eastern population of the monarch butterfly to 225 million butterflies occupying an area of approximately 15 acres (6 hectares) in the overwintering grounds in Mexico."

There are currently over 1,350 animal species[27] listed as endangered or threatened under the Endangered Species Act of 1973, which include some well-known species such as sea turtles, wolves, and bears. Approximately 76 peer-reviewed articles in economics have studied endangered species (Pandit, et al., 2015). Often economists and the public focus on "charismatic species." Charismatic species are usually a large, easily identifiable species that have widespread popular appeal and often used by achieve broader environmental

[27] Statistics generated on 2017-02-08 from http://ecos.fws.gov/ecp0/reports/box-score-report

goals (Ducarme, et al., 2013).[28] In this study, we define charismatic as being well-known by the public *and* being TER.[29]

Loomis and White (1996) analyzed 18 TER species' economic value collected via stated preference approaches, and a majority of these were oriented towards iconic birds or mammals.[30] Brown and Shogren (1998) later commented that the average value of the 18 species in total was about $1000 per household, and if it were aggregated across all households, it would represent "1% of the 1995 U.S. Gross National Product, for less than 2% of all threatened and endangered species," values that many would deem "suspiciously high."[31] In reality, the aggregated WTP was $953, but was made up of studies that reported annual WTP ($362) as well as lump-sum WTP ($591) for the various species. Annuitizing the lump-sum values generates an annual WTP of $47.42[32], so that a more accurate depiction of annual WTP is $409.42, rather than the originally quoted $1000. All the same, this represents WTP equal to $644 in 2016 dollars, which many may still guess to be an overestimate.

---

[28] Verissimo et al. (2009) even identified which bird species were the strongest candidates to use in public campaigns.

[29] A widely-accepted definition 'charisma' does not seem to exist, and has been a point of controversy for some time (see Metrick and Weitzman, 1996). For instance, Walpole and Leader-Williams (2002) state that a charismatic species does not have to be endangered, but species must have a compromised conservation status in Clucas et al. (2008).

[30] Later expanded by Richardson and Loomis (2009) to 67 observations from 31 studies.

[31] To their credit, Loomis and White specifically state their purpose "is not to provide such aggregate estimates."

[32] Assuming a 5% discount rate and 20-year annuity.

One explanation of these seemingly high estimates could be that charismatic species represent broader support for biodiversity, not just for the species itself. Some portion of these values represent the additional WTP for those particular species' charisma, value beyond the normal economic values such as existence, option, or bequest.

Because valuation of charismatic species often relies on stated preference methods, another explanation of such inflated values is HB. With a few exceptions, underlying explanations of the persistence of HB are rare. Given the exceptionally high WTP estimates documented in previous works, we investigate the extent of additional HB due to charisma. In this case, the monarch butterfly has received a high amount of publicity and national attention to its plight, and could be considered a charismatic species useful to identify HB due to charisma.[33]

An ideal identification strategy would implement a split-sample design in which one group values a charismatic species, and the other values an identical, non-charismatic species. Because charismatic species are often megafauna (large, iconic mammals such as polar bears, lions, whales, etc.), formulating this type of design is extremely difficult using two real species because of the difficulty to identify a non-charismatic counterfactual. For monarchs, this question can be answered because of the existence of the viceroy butterfly. The viceroy and monarch butterfly are visibly nearly identical and have near identical ranges

[33] In their review of economic studies of endangered species, Pandit et al. (2015) classify Monarchs as a charismatic species.

across North America, especially in the region pertaining to the study respondents.[34]

By comparing how individuals value monarchs and viceroys differently, the difference represents the charisma of the monarch. While monarchs, as insects, are not a perfect representation of previously studied charismatic species, this difference can provide one explanation of HB within the context of charismatic/TER species. Further, it contributes to the dearth of valuation literature on insects.

In summary, through a choice experiment, our experimental design enables us to address the following questions:

1) What is the extent of hypothetical bias for monarch and viceroy butterflies?
2) What is the extent of hypothetical bias with respect to distance decay for monarchs and viceroy butterflies?
3) As a measure of charisma, what is the additional WTP associated with monarchs compared to viceroys?

[34] The viceroy is slightly smaller and has one subtle difference in wing pattern. This similarity is known as Müllerian mimicry (Ritland and Brower, 1991), when two species mutually benefit from displaying the same warning signal. Focus groups and pre-test of our survey suggest that respondents cannot differentiate these two butterfly species beyond a random guess.

We combine the information to generate a rough correction for an estimated real value of the 18 TER species from Loomis and White (1996) as well as other implications.

3.4 Study Design

3.4.1 Choice Experiment and Survey Design

To answer our research questions, we utilized a 2x2 experimental design in conjunction with a CE. As a split-sample design, each participant in our study could be in one of four treatments: a real or hypothetical valuation and valuing either monarch butterflies or viceroy butterflies. The CE was designed with the goal of understanding values of butterfly conservation among participants from the city of Lexington, Kentucky. The CE's attributes and corresponding levels are described in **Table 3-1**.

The good presented to respondents was a donation to purchase and install plants that support butterfly conservation. This good was chosen for a number of reasons. The dearth of milkweeds and nectar plants for monarchs along their migration routes and summer breeding grounds is one of the primary theories for the monarch's dramatic decline (Flockhart, et al., 2015, Pleasants and Oberhauser, 2013).[35] Installing plants for the monarch's benefit is a widely-

[35] Inamine et al. (2016) demonstrate that this belief is not held universally by all entomologists.

accepted mechanism to support monarch conservation. Additionally, installing plants has the benefit of being tangible and divisible.

Participants were told that all donations go towards the purchase and installation of plant seedlings, each at a cost of $1. The cost information was obtained and confirmed through checking multiple nurseries in or around the city. Similar to Ready, et al. (2010), this means the good is quasi-public, additional benefit to butterflies is only provided if the respondent donates, mitigating free-riding behavior. Given this information, a donation towards the installation of additional plants that support butterflies was chosen as the most credible good.[36] Upon multiple focus group and pilot testing, the potential donation between $1 and $10 is deemed reasonable.

Three non-payment vehicle attributes were part of the CE for installing plants: the location, site accessibility, and designation as a Waystation. The three locations, Paducah (McCracken County), Elizabethtown (Hardin County), Lexington (Fayette County), were deliberately chosen. All three are among the largest of Kentucky's statistical areas. Their separation is rather linear, avoiding the potential of directional effects as observed in Schaafsma, et al. (2012). The driving time, between Paducah and Elizabethtown and between Elizabethtown and Lexington is 2.5 and 1.5 hours, respectively. By keeping the benefit of

[36] Other mechanisms may be possible. For example, one alternative is to donate to support monarch overwintering sites in Mexico. This has the disadvantage of being more abstract, create potential free-riding, generate potential geopolitical distortions, and most importantly, is inapplicable to viceroy butterflies.

donations within the state, it reduces the chance of potential geopolitical threshold effects which can be confound with potential distance decay effects (Johnston and Duke, 2009, Rolfe and Windle, 2012, Van Bueren and Bennett, 2004). Distance was stressed to respondents in a number of ways. In the CE instructions, respondents saw a map of Kentucky highlighting each of the 3 locations to visually reinforce the distance of Elizabethtown and Paducah from Lexington as well as the estimated drive time to each from Lexington. These locations for installing plants that support butterfly conservation is the primary mechanism for testing for distance decay and will be explained in detail below.

The next attribute is the accessibility of each butterfly restoration site, such that a respondent could or could not physically visit and/or see a site. This is similar to Johnston and Ramachandran (2014). Access could be considered a measure of the respondent's option value.

Lastly, each location could become a certified "Monarch Waystation," which included the installation of a corresponding sign, and described as supporting the conservation of many butterfly species.[37]

Respondents were informed that the Waystation certification and sign installation occur after a habitat is created, which means that the benefits to butterflies is independent of whether a habitat is a certified Waystation. Conversely, the designation and sign increase each a habitat's outreach and

[37]To be truthful, respondents were informed there was no viceroy-specific Waystation program, but that Waystations promote butterfly conservation of many species, and listed some examples of other species.

educational ability to the public. Each respondent answered six choice sets, with a sample choice set featured in **Figure 3-1**.[38]

Figure 3-1 also shows that the CE used a repeated binary choice format, a first alternative with varying attribute levels, and a second opt-out alternative that provided no support for butterfly conservation nor any payment by the respondent. This binary elicitation mechanism was chosen because a single binary choice can be incentive-compatible (Carson, et al., 2014, Vossler, et al., 2012) under certain conditions. These conditions are that respondents care about the outcome, that payment is enforceable, elicitation is a yes/no vote for a single project, and that likelihood of provision increases with proportion of yes votes.[39]

Further, Vossler, et al. (2012) show that with some additional conditions, a sequence of binary choices can still maintain incentive-compatibility. These are: that only one of the series of binary choice sets will be implemented, that provision in each choice set is independent of decisions in other choice sets, and that the characteristics in the choice set exactly correspond to the policy implemented and no other policy.

[38] Image of the Waystation sign was used with permission from Monarch Watch.

[39] A single dichotomous-choice elicitation can be considered a specific form of a voting-style elicitation such that it is a referendum determined by one person, in which the person's vote entirely determines provision (Answering no means no provision nor payment with 100% certainty, and answering yes means provision and payment occurs with 100% certainty).

Some of these assertions may be rather strong in a field survey setting, but we will describe steps taken below to make such assertions more plausible. Correspondingly, we avoid a multinomial CE to circumvent the considerable doubt of its incentive-compatibility, formalized by the Gibbard-Satterthwaite theorem (Gibbard, 1973, Satterthwaite, 1975). The CE's design allow for tests of distance decay for values of monarch and viceroy butterflies.

To implement the CE, we use a full factorial design, using 36 two-alternative choice sets. Each respondent participated in one of six groups of choice sets, and each group contained six choice sets. After completing their choice sets, respondents assigned to a treatment requiring actual payment rolled a 6-sided die to determine which of the choice sets would be binding. If their answer in the binding choice set was to donate, the respondent immediately placed the corresponding amount in a secured lock-box. Afterwards, they continued the survey until completion.

With respect to identifying differences in the value of monarch and viceroy butterflies, respondents read a brief description of only one butterfly species.[40] To ensure reading comprehension, each respondent answered several True-False questions on whether their butterfly's range included the entire state of

[40] Note that the description did *not* include a picture of the specific butterfly. Specifically, focus groups and pilot testing revealed that even though respondents read a description and saw a picture of a viceroy, they frequently associated the picture with a monarch butterfly anyway. To avoid confusion between the two butterflies, we choose not provide a picture of either butterfly.

Kentucky and whether their butterfly was considered a vulnerable species. Further, if respondents provided an incorrect answer, a brief message reminded respondents of the correct answer. This approach of reminding respondents improves respondent cognition of the range and status of each butterfly species. This design coupled with the similarity of the viceroy and monarch butterflies means any difference in values between the two species will likely be attributed to the charisma of the monarch butterfly, both in hypothetical and real valuations as well as associated HB. Beyond the CE, the survey included a variety of other questions such as attitude towards conservation, knowledge and interest in butterflies, as well as standard demographic queries. Our central hypotheses are:

Hypothetical Bias (H1)

$H1_{0:}$ hypothetical WTP is less than or equal to real WTP for both monarch and viceroy butterflies;

$H1_A$: hypothetical WTP is greater than real WTP for at least one of the butterfly species, i.e. the presence of HB.

Distance Decay (H2)

$H2_0$: WTP for a Lexington site is less than or equal to the WTP for an Elizabethtown site, and/or WTP for an Elizabethtown site is less than or equal to the WTP for a Paducah site, and/or WTP for a Lexington side is less than or equal to the WTP for a Paducah site, regardless of whether the treatment is real, hypothetical, monarch, or viceroy;

$H2_A$: Distance decay holds between at least some of the three locations and in one of the treatments.

Distance Decay HB (H3)

$H3_0$: Conditional on evidence to support HB ($H1_A$), the extent of HB is equal for all three locations regardless of butterfly species;

$H3_A$: HB is not equal across locations for at least one butterfly species.

Charisma (H4)

$H4_0$: WTP for monarchs is less than or equal to that of viceroys for both the real and the hypothetical comparisons and regardless of locations;

$H4_A$: WTP for monarchs exceeds WTP for viceroys in at least one of the treatments in one of the locations.

Charisma HB (H5)

$H5_0$: Conditional on evidence to support HB ($H1_A$) and charisma ($H4_A$), the extent of HB for monarchs is less than or equal to that for viceroys;

$H5_A$: HB for monarch butterflies is greater than HB for viceroy butterflies.

3.4.2 Field Survey Implementation

We implement a field survey using the CE and experimental design described above. All respondents were from the single metropolitan community, Lexington, in the state of Kentucky. This has the inherent benefit of mitigating differences in value due to proximity to the resource. Similarly, because of both species ubiquity throughout the state, our analysis and sample focuses almost exclusively on distance decay and avoids spatial cold or hotspots (Johnston and Ramachandran, 2014) when sampling over a larger scale.

Prior to launch, the survey went through multiple rounds of refinement based on four focus groups as well as a pilot survey. Surveys were completed on an internet-connected tablet, which allowed for treatment randomization and enhanced audio-visual communication with respondents. Specifically, after reading a description of their respective butterfly species, respondents watched a one-minute video for the CE instructions.

It is typical to provide text-based instructions, but focus group feedback demonstrated that communication via video instructions along with an example choice set in the video improved respondent comprehension when completing the CE choice sets. Respondents assigned to the real payment treatment group watched a slightly longer video in order to explain how the roll of a die would be used to determine the binding choice set. Additionally, a true-false question appeared immediately after the video with a statement to reaffirm that the respondent understood they would be expected to pay based on the roll of the die.

Surveys were collected during May, June, and July of 2016 on 51 occasions at 35 unique locations or events and occurred at least twice every day of the week at various times (e.g. morning, afternoon, and evening) throughout Lexington. Collection occurred as early as 8am to as late as 9pm, but responses tend to come from weekday afternoons/evenings and weekends. While each of the survey collection sites were outdoors, which is common for an environmental and resource valuation study, they did not necessarily focus on outdoor enthusiasts. For example, surveys were collected at a county fair, at a movie in

the park, at playgrounds, at sports events, and at jazz festivals. This makes it possible for the sample to be qualitatively similar to the general Lexington population, though we do not claim it is representative of the broader US population.

During each occasion, the same equipment and promotional material was used to provide a consistent visual presentation. To reduce interviewer bias, one survey enumerator was present at all events as well as an assistant enumerator, which rotated among five other individuals.

Once a potential respondent agreed to participate, they were seated in front of a tablet to begin the process. Prior to starting the survey itself, each respondent completed a separate exercise to earn $10. This is to allow the respondents to treat the money as earned instead of windfall/house money, the latter of which may distort WTP (Clark, 2002, Loureiro, et al., 2003). To match the potential $1, $5, and $10 payments in the real CE, both hypothetical and real respondents received five $1 bills and one $5 bill. To mitigate protests, real respondents were notified at the beginning the survey that they would have a chance, but would not be obligated to make a real donation during the survey.

3.5 Model Specification

Discrete choice models are based upon Random Utility Theory, which describes a person's utility from a particular good being composed of observable and unobservable components (McFadden, 1973). Equation 1 shows that

individual *i* derives utility from selecting alternative *j* in choice set *t* with observable attributes X_{ijt} and the payment variable c_{ijt} and an unobservable component, $\boldsymbol{\varepsilon}$:

$$\boldsymbol{U_{ijt} = -(\alpha_i/k_i)\, c_{ijt} + (\beta_i/k_i)' X_{ijt} + \varepsilon_{ijt}} \quad (1)$$

Among the coefficients to be estimated, α represents the effect of change in price, while the vector for β yields the estimated effect of various attributes on their choice. The coefficients are indexed by *i* to show that the effect of attributes varies across individuals, one of the primary advantages of conducting a mixed logit model based on (1).

The above specifications represent a model in parameter space. In typical parameter space models, the scale parameter, k_i, is inherent to but not separately identifiable in the model, and is assumed to be fixed, such that the unobservable component's variance is equal across respondents (i.e. homoscedasticity). This issue of scale has two important implications: 1) comparing coefficient estimates across samples is inappropriate due to scale differences, and 2) that the variability in unobserved utility is the same for all respondents, which can potentially bias other coefficient estimates in the model.[41] If $\lambda_i = (\alpha_i/k_i)$ and $h_i = (\beta_i/k_i)$, then WTP, ω_i, is simply h_i/λ_i, which eliminates the scale issue.

[41] Train and Weeks (2005) mention other disadvantages of parameter-space models are that the price coefficient is usually fixed across respondents, implying a constant

We address these parameter-space issues by modelling choices in WTP-space as in equation 2. Train and Weeks (2005) demonstrate its equivalence to parameter-space.

$$\boldsymbol{U_{ijt}} = -\boldsymbol{\lambda_i}\left[\, \boldsymbol{c_{ijt}} + \boldsymbol{\omega_i'X_{ijt}}\right] + \boldsymbol{\varepsilon_{ijt}} \tag{2}$$

Equation 3 reflects WTP-space in our application with the omission of subscript *i*. As such, ω_l reflects WTP estimates per attribute for the reference group made up of Real-Viceroy respondents and δ_m represents the change in WTP for the various treatment groups (Hypothetical-Viceroy, Real-Monarch, and Hypothetical-Monarch) relative to the reference group.

marginal utility of income. If a distribution is assumed, then the associated WTP, usually the ratio of a normally-distributed attribute coefficient to a log-normally-distributed payment vehicle coefficient, has undefined moments. Secondly, assuming independent parameter-space estimates of attributes implies correlated WTP across attributes.

$$
\begin{aligned}
U_{jt} = -\lambda\big(&\boldsymbol{Donation\ Amount}_{jt} + \omega_1 \boldsymbol{OptOut} + \omega_2 \boldsymbol{Elizabethtown}_{jt} \\
&+ \omega_3 \boldsymbol{Lexington}_{jt} + \omega_4 \boldsymbol{Public}_{jt} + \omega_5 \boldsymbol{Waystation}_{jt}\big) \\
&+ \sum_m \delta_{m1}(\boldsymbol{OptOut} * \boldsymbol{Treatment}_m) \\
&+ \sum_m \delta_{m2}(\boldsymbol{Elizabethtown} * \boldsymbol{Treatment}_m) \\
&+ \sum_m \delta_{m3}(\boldsymbol{Lexington} * \boldsymbol{Treatment}_m) \\
&+ \sum_m \delta_{m4}(\boldsymbol{Public} * \boldsymbol{Treatment}_m) \\
&+ \sum_m \delta_{m5}(\boldsymbol{Waystation} * \boldsymbol{Treatment}_m)) + \boldsymbol{e}_{jt}
\end{aligned}
\tag{3}
$$

By having coefficients directly represent WTP, the issue of scale is removed. As seen in equation 3, it allows for data from different treatments to be pooled and directly test for differences by including interaction terms. Further, modelling in WTP space allows for scale heterogeneity across respondents, which is represented by the standard deviation of the payment vehicle. [42] Lastly, WTP-space assumes a distribution of WTP itself, such that the ratio is assumed to be normally distributed, rather than the problems of assuming a distribution for the numerator and denominator (see Carson and Czajkowski (2013)).

[42] Allowing for scale heterogeneity is also possible in parameter-space by using generalized multinomial logit (gmnl) models (Fiebig et al. 2010). In fact, Greene and Hensher (2010) show that WTP Space is a special case of gmnl.

Formal comparisons of WTP space and parameter space remain relatively sparse. Nevertheless, several cases show that WTP space models produce more reasonable estimates of the distribution of WTP versus parameter space models (Hole and Kolstad, 2012, Scarpa, et al., 2008, Train and Weeks, 2005).

The opt-out constant represents a choice not to donate in a particular situation. It usually represents the disutility of being unable to consume the offered good with the base level of the various attributes. In our case, this is the installation of plants in Paducah, KY in a private location without the waystation designation (presumably the least valuable alternative possible). We utilize a mixed logit model assuming that WTP for Opt-Out, Elizabethtown, Lexington, and Public are heterogeneous following a normal distribution while the Waystation attribute remains fixed. The Waystation attribute is specified with a non-random WTP measure because in various trial analyses, the standard deviation of this WTP measure is always insignificant. We use 250 Halton draws in WTP space. We rely on the delta method for a number of post-estimation comparisons of model results.

3.6 Results

3.6.1 Sample Description

In total, 789 useable responses were collected in the field survey. Select socioeconomic characteristics of the sample respondents, both per treatment and collectively, are presented in **Table 3-2**.

First, based on demographic information, no significant differences exist in demographic characteristics across the four treatment groups. While the treatments are statistically similar, taken together, the sample is not perfectly representative of the community. It resembles the community reasonably with respect to age and gender, but dissimilar with respect to children and educational levels.

Across all treatments, a total of 141 respondents chose not to donate in all six of their choice sets. A follow-up question revealed that 55 were (34 from hypothetical and 21 from real) protest respondents, allowing for an analysis based on a total of 734 respondents.

3.6.2 Model Results

Mixed logit WTP-space model results including treatment interactions are presented in **Table 3-3**. We first focus on the results of the baseline, Real-Viceroy respondents. Individual coefficient estimates follow expectations. Scenarios with higher requested donations are significantly less likely to be chosen, and publicly accessible locations are more likely to be chosen. We do not find evidence that the Waystation designation and associated sign as being significant in affecting respondent choice. Lastly, we observe some evidence of distance decay in that the WTP for viceroy conservation is greater in Elizabethtown and Lexington, discussed in more detail below. Since the focus of this study is on HB and distance decay, in the following discussion, we base our

interpretation on conservation sites located on private land (variable Public = 0) without a monarch Waystation designation or sign (variable Waystation = 0).

While it may seem peculiar to observe a significant and positive WTP for the opt-out alternative, this result is unsurprising in the current context. Because the donation is ultimately for the installation of plants for a non-endangered butterfly species in a distant location (not a representation of the species itself), it is reasonable to expect that, for many people, the utility of keeping their money for other activities would exceed the utility of a donation. In this case, viceroy respondents receive positive utility equal to $4.81 to avoid making a donation. Equivalently, the dollar value of disutility from forcing a respondent to support viceroys is $ -4.81. Since monarchs are well known and potentially endangered, we would expect and find that the disutility to support plants for its conservation to be smaller, equal to $3.28 ($4.81-$1.53) in the Real-Monarch treatment and $.08 ($4.81-$4.73) in the Hypothetical-Monarch treatment with the latter being insignificant from zero.

From the results of the standard deviations, we observe significant differences across individuals for each of the attributes. The significance of the donation amount means that there is significant scale heterogeneity across respondents. Furthermore, the standard deviations are roughly twice as large as their corresponding point estimates of WTP. This suggests an extremely wide range of values associated with butterfly conservation.

To begin our comparison across treatments, we first consider the extent of HB for viceroys and monarchs. If HB exists, WTP to opt out in hypothetical

treatments will be closer to 0, indicating of smaller penchant to opt-out, all else equal. Because the coefficient of opt-out in Hypothetical-Viceroy is not significant, we cannot reject the null hypothesis, $H1_0$, that there is no HB in the opt-out for real versus hypothetical viceroy respondents.

To determine HB for monarchs, we observe that the opt-out WTP for Real-Monarch ($3.28=$4.81-$1.53) is significantly greater than Hypothetical-Monarch ($.08=$4.81-$4.73) (p=.04). This means there is evidence of HB for monarchs, supporting $H1_A$. Therefore, for the baseline location (i.e., Paducah), we have evidence of HB for monarchs, but not for viceroys.

Next, we consider distance decay. If distance decay exists, then we would expect that the coefficients of Elizabethtown and Lexington to be positive, with Lexington being larger in magnitude compared to Elizabethtown. In the Real-Viceroy treatment, compared to the reference location of Paducah, Lexington is statistically significant, with respondent WTP equal to $4.36, but Elizabethtown is not significantly different. This supports $H2_A$ and demonstrates distance decay for Real-Viceroy. Importantly, rather than linear decay, it is a sharp decline with relatively little value outside of Lexington.

We reach similar conclusions in support of $H2_A$ for the Real-Monarch treatment. Since neither the Elizabethtown nor Lexington interaction coefficients are significant, the combined effect is still that Lexington conservation sites are associated with a larger value than sites in the other two locations. This again implies a similar distance decay pattern as for Real-Viceroy.

In order to test H3, we examine WTP for the hypothetical treatments of monarch and viceroy at each location. Coupling this with real WTP information can let us determine HB in each location. Recall that the opt-out of Hypothetical-Viceroy was not significantly different to that of Real-Viceroy, suggesting no HB in the opt-out for viceroys. Once location is included, we find marginal evidence that WTP is higher for Hypothetical-Viceroy in Lexington. This implies that there is some evidence of HB for viceroys in Lexington, but because the Hypothetical-Viceroy interactions for opt-out (representing Paducah) and Elizabethtown are not significant, there is no evidence to suggest HB with respect to distance decay for viceroys, supporting $H3_0$. For monarchs, the significance of the opt-out for Hypothetical-Monarch is especially important. It indicates that, even while the WTP for Elizabethtown and Lexington are not significantly different from each other in the two monarch treatments, hypothetical WTP values exceeds real WTP in all three locations, which in turn means there is HB even in locations that are more distant. This is evidence of $H3_A$ for monarchs.

Figure 3-2 displays the WTP for in each location for all four treatments using all estimated location coefficients, regardless of statistical significance. To facilitate comprehension, we use the negative of the opt-out coefficients, again representing the value if forced to donate. This makes it clear that outside of Lexington, the WTP to support butterfly conservation is less than or equal to 0. In Lexington, only Hypothetical-Monarch and Hypothetical-Viceroy are significantly greater than 0 (both p-values<.01).

Another important comparison is to identify whether there is a premium for charisma received by monarchs relative to viceroy butterflies (H4), and if HB affects this premium (H5). If there is charisma, we would expect the WTP to opt-out to be closer to 0 for monarchs, in other words, the disutility of a forced donation should be smaller for monarchs.

Establishing the value of charisma can occur based on two comparisons of monarchs and viceroys, either real or hypothetical WTP. Based on the non-significance of the Real-Monarch opt-out coefficient compared to that of the Real-Viceroy, we observe no real charisma premium for monarchs and no evidence to reject $H4_0$. In the second comparison of hypothetical treatments, we find evidence of a charisma premium for monarchs compared to viceroys. The hypothetical WTP to opt-out for monarchs is $.08 ($4.81-$4.73), while the hypothetical WTP to opt-out for viceroys equals $4.63 ($4.81-$.18), and the two are significantly different ($p<.001$), which supports $H4_A$.

These results provide at least initial evidence that charisma has a considerable effect on hypothetical WTP, but not on real WTP, therefore, HB may be more pronounced for a charismatic species versus their non-charismatic counterparts. This finding particularly calls into question of the previous analysis of the value of charismatic species based on hypothetical surveys. Using the results from **Table 3-3** to test H5 on the difference in HB for monarchs and viceroys. We find a significant difference ($p=.031$) in the HB of viceroys ($4.63/$4.81) to the HB of monarchs ($.08/$3.28) in Paducah. We attribute this difference in HB to the charisma effect, evidence to support $H5_A$. A similar

analysis can be completed for Elizabethtown or Lexington, but in both cases, there was no significant difference (p=.165 and p=.211, respectively).

Given the evidence of additional HB for monarchs, and using hypothetical and real WTP for monarchs and viceroys, we have enough information to calculate the extent of HB for the charismatic species and how much of the HB is due to charisma, as seen in **Table 3-4**. In this calculation, we use the WTP estimates based on Lexington: while not significant, this approach facilitates interpretation and is most appropriate due its proximity to the respondents. A similar conclusion is reached using WTP in Paducah of the four treatments (also reported in **Table 3-4**).

In this case, we take the negative of the opt-out constant because making a donation to Lexington inherently means that the respondent faces the disutility of the opt-out combined with the utility of donating to Lexington. We observe that the difference between hypothetical and real WTP for viceroys and monarchs is $2.84 and $4.17, respectively. This means that monarchs have about $1.33 more HB, or about 31.9% of the $4.17 difference between hypothetical and real WTP for monarchs.

Based on our estimates and HB reduction for charismatic species, it may be appropriate to reduce the total value of TER in Loomis and White (1996). We observe that the difference between hypothetical and real WTP for monarch conservation in Lexington is $4.17, or approximately 83.1% of hypothetical WTP. If we apply this reduction to the $409 for the 18 species from Loomis and White (1996) calculated earlier, the estimated real WTP decreases by $301 to $69 ((1-

.831)*409). In 2016 dollars, this is a correction from $644 to $109. While this a rough estimate, an average of $6 per species is unlikely to draw the attention of economists as “suspiciously high.” Further, because of our previous model results, we estimate that 31.9% of the $301 reduction is due to HB from charisma ($95.89) and the remaining 68.1% ($204.70) is from typical HB. For comparison, an even larger proportion (94.4%) of HB for monarchs in Paducah is attributable to the charisma effect. Caution is warranted in this correction though because we are using our data of butterflies to suggest deflated values for many types of charismatic megafauna.

As an additional check, we present the WTP per attribute for each treatment based on mixed-logit parameter space in **Table 3-5**. These results are based on the Krinsky-Robb approach using 5,000 permutations. Results are largely consistent in terms of sign, significance, and magnitude of the parameters.

3.7 Discussion

We investigate the extent of HB with respect to distance decay and charismatic species through a valuation of butterfly conservation. Based on this analysis, we find a number of results.

First, there is distance decay in WTP for both monarch and viceroy conservation; people prefer to support conservation in their own community compared to a more distant one. Given the ubiquity of monarchs and viceroys

throughout the state, the sense of ownership to the resource seems the most probable of Rolfe and Windle's (2012) four principal reasons for distance decay.

Second, when we compare hypothetical and real WTP across locations, we observe HB in distance decay for monarchs, but no such HB in distance decay for viceroys. There is still some evidence of HB for viceroy though because hypothetical WTP is greater than real WTP in Lexington.

Third, we find that WTP for monarchs exceeds viceroys in the hypothetical treatment, but not in the real treatment, it suggests there is a hypothetical charisma effect. On the other hand, WTP is equal to support monarchs and viceroys in the real treatments, indicating that the two species are valued equally and no evidence of a real charisma effect. Because the two butterflies are so similar, this means there is additional HB for monarchs compared to viceroys, evidence that the additional HB is due to a charisma effect.

Interestingly, many conservation organization use charismatic species as 'flagship species' as a way to improve fundraising and campaign effectiveness (Ducarme, et al., 2013). Our results show that, at least with respect to monarch butterflies, only hypothetical donations are likely to see a flagship premium, and the real benefits of a flagship species are much lower than what may be presumed. Equivalently, in most previous studies that used hypothetical survey to elicit public WTP for symbolic species, the suggested WTP may due to HB as well as charisma effect. If one uses a real WTP eliciting technique, it may reduce typical HB as well as HB from the charisma effect.

Some qualifications of the research design exist. First, our results of distance decay use locations within one state for a species that is nationally present. Similarly, our design varies the location where the conservation effort occurs while the respondent's location remains fixed. Most distant decay valuation studies do the opposite, focusing on a resource at a fixed location and sampling respondents at varying locations.

Additionally, monarch butterflies, even if categorized as a charismatic species, are charismatic insects, which are not equivalent to charismatic mammals. Our finding that WTP reduction of $831 for the 18 species in Loomis and White, $265 (31.9%) of it stems from the charisma effect may be an underestimate. Compared to insects, mammals are relatively 'more charismatic', so would likely have a larger proportion of their inflated WTP due to the charismatic effect.

The question remains, why are charismatic species more likely to have HB? One explanation of HB in the context of species conservation that may be especially important is social desirability bias. Because of a charismatic species' ubiquity, people generally know that the "correct" answer in society is to show support, financial or otherwise, easily achieved in a purely hypothetical survey.

Table 3-1 Choice Experiment Attributes and Levels

Attribute	Description	Levels
Location	Potential sites in Kentucky to install butterfly plants	1. Lexington (Fayette County) 2. Elizabethtown (Hardin County): 85 miles away from Lexington (1.5 hour drive) 3. Paducah (McCracken County)*: 350 miles away from Lexington (4 hour drive)
Accessibility	Public's ability to visit site	1. Closed*: habitat inaccessible nor viewable by the public, such as a private farmland 2. Open: habitat accessible and viewable by the public, such as public parks
Waystation	Inclusion in national waystation program	1. Certified: Waystation is certified and Waystation Sign is installed. 2. Not Certified*: Habitat is not a certified Waystation nor is a Waystation Sign installed.
Donation	Amount of money to support butterfly plants	$1, $5, $10

* indicates reference category in CE

T**able 3-2** Select Sample Characteristics (all entries are percentage measures)

Variable	Population[1]	Total Sample	Real Monarch	Hypo Monarch	Real Viceroy	Hypo Viceroy
N[2]		734	147	223	147	217
Age						
18-24	18.6	17.3	15.1	19.3	15.7	18.0
25-34	19.4	28.5	31.5	31.8	27.2	24.0
35-44	16.8	23.5	20.6	19.3	26.5	27.7
45-54	16.0	14.7	16.4	15.3	13.6	13.8
55-64	14.6	10.8	11.0	11.7	10.2	10.1
65+	14.7	5.2	5.5	2.7	6.8	6.5
Chi-2(15)=13.6, p-value=.56[3]						
Male	48.6	43.3	46.9	39.9	40.4	46.3
Female	51.4	56.7	53.1	60.1	59.6	53.7
Chi-2 (3)= 3.1, p-value=.37						
Education						
High school or less	30.0	21.1	21.8	20.4	27.2	17.1
Some college	27.4	24.8	22.5	23.1	27.9	25.9
Bachelor's degree	23.6	27.9	27.9	29.0	23.1	30.1
Graduate/professional	17.0	26.3	27.9	27.6	21.8	26.9
Chi-2 (9)=8.7, p-value=.47						
White	75.6	71.5	70.1	73.1	70.8	71.4
Black/African American	14.4	14.2	13.6	13.9	12.9	15.7
Asian	3.6	2.9	4.1	1.8	1.4	4.2
Chi-2 (6)=4.4, p-value=.62						
Minors at home	28.9	46.1	43.8	48.9	48.3	43.3
Chi-2 (3)=2.0, p-value=.58						
Single, never married	38.8	33.1	30.6	35.9	34	31.3
Married	41.1	53.4	53.1	50.7	50.3	58.5
Chi-2 (3)=2.1, p-value=.54						
Median Income	$47968	$42,500*	$42,500*	$42,500*	$42,500*	$62,500*

[1] Based on 2015 ACS 1-year
[2] Based on sample of non-protest respondents
[3] Chi-square tests are used to test for differences across the four treatment groups
*Value calculated using midpoint of responses

Table 3-3 WTP-Space Model Results for Butterfly Valuation[1]

N=734	**Baseline for Real-Viceroy WTP**	**Baseline plus Hypothetical-Viceroy WTP**	**Baseline plus Real-Monarch WTP**	**Baseline plus Hypothetical-Monarch WTP**
Ln(Donation)	-0.77***			
	(0.17)			
Mean WTP				
Opt-Out	4.81***	-0.18	-1.53	-4.73***
	(1.33)	(1.55)	(1.83)	(1.65)
Elizabethtown	1.67	-0.5	-1.53	-1.4
	(1.12)	(1.43)	(1.51)	(1.38)
Lexington	4.36***	2.66*	-0.23	0.74
	(1.22)	(1.55)	(1.7)	(1.48)
Public	1.76**	1.77	1.43	1.85*
	(0.84)	(1.08)	(1.16)	(1.07)
Waystation	-1.40	2.12*	1.54	1.84
	(0.94)	(1.15)	(1.25)	(1.14)
Standard Dev.				
Ln(Donation)	0.72***			
	(0.23)			
Opt-Out	9.71***			
	(0.64)			
Elizabethtown	3.81***			
	(0.73)			
Lexington	7.3***			
	(0.65)			
Public	4.46***			
	(0.53)			

[1] Standard errors reported in parentheses; ***, **, and * indicate statistical significance at the p-value<.01, <.05, and <.1, respectively.

Table 3-4 Estimated WTP among respondents for butterfly conservation based on results of Table 3-3

	Real	Hypothetical	Difference	Difference as a % of Hypothetical WTP
Lexington-Viceroy	-$.45	$2.40	$2.84	NA[1]
Lexington-Monarch	$.85	$5.03	$4.17	83.1%
		Difference=	$1.33 (31.9% of $4.17)	
Paducah-Viceroy	-4.81	-4.63	$0.18	NA
Paducah-Monarch	-3.28	-.08	$3.20	NA
		Difference=	$3.02 (94.4% of $3.20)	

[1] NA indicates that this number is uninterpretable in the conventional sense of Hypothetical Bias for WTP

Table 3-5 Mixed Logit Parameter Space WTP for Viceroy and Monarch Butterflies[1]

	Real-Viceroy WTP	Hypothetical-Viceroy WTP	Real-Monarch WTP	Hypothetical-Monarch WTP	Poe Test[2]
Opt-Out	1.25	1.63	.71	.03	C,D
Elizabethtown[NS]	1.05	1.07	-.14	-.04	
Lexington	4.71	7.53	2.95	5.48	A,B
Public	2.07	3.84	2.27	4.23	A,B
Waystation	-1.23	.82	-.20	.49	A

[1]Based on mixed-logit parameter-space model results.
[2]A, B, C, and D indicates a significant difference ($p<.1$) in WTP between real and hypothetical Viceroy, real and hypothetical monarch, real viceroy and real monarch, and hypothetical viceroy and hypothetical monarch, respectively.
[NS] Indicates underlying parameter estimates were not significant.

Option A	Option B
$5 donation  Paducah Public 	$0 donation No additional contribution to butterfly conservation

Figure 3-1 Example Choice Set

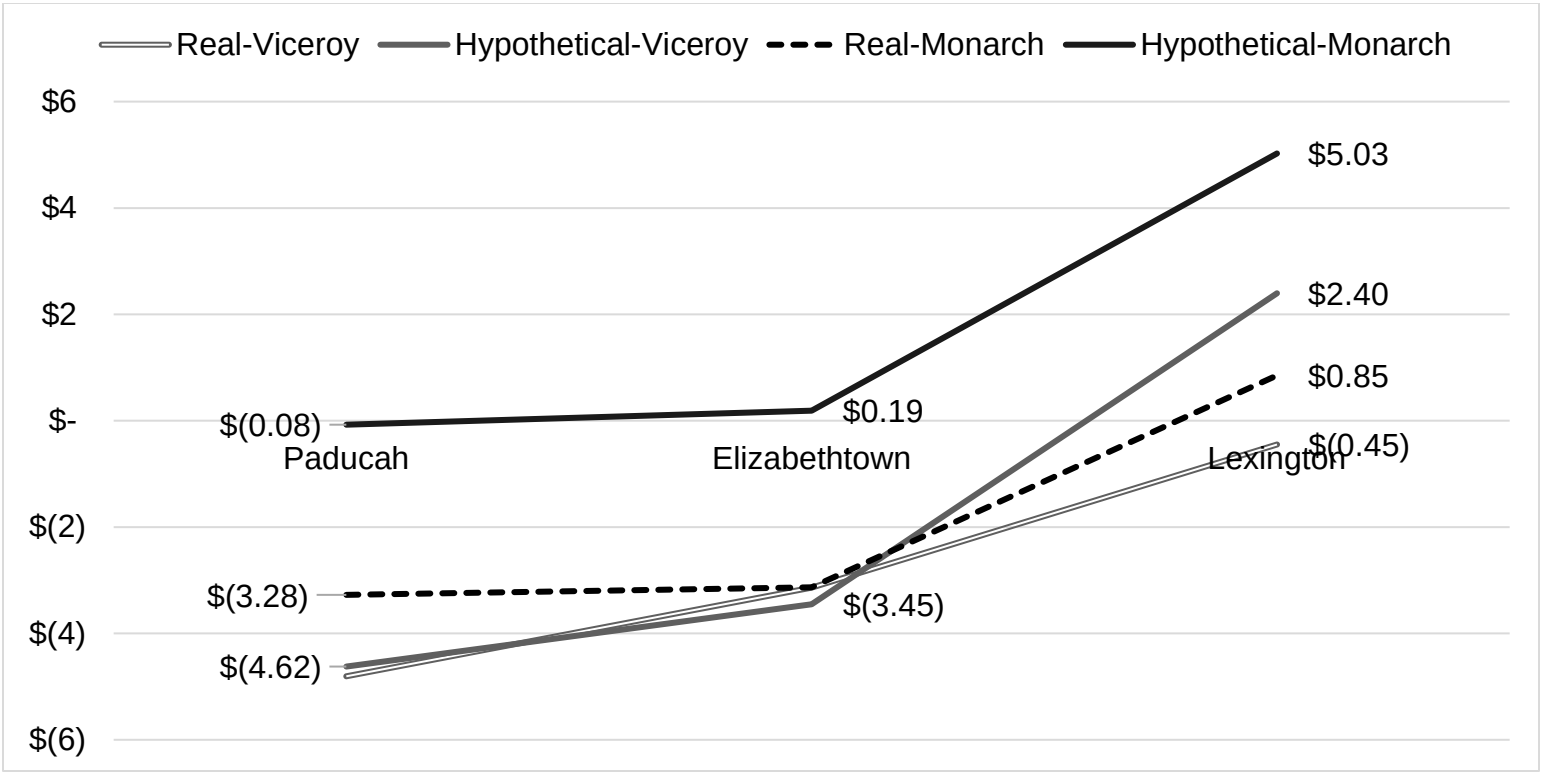

Figure 3-2 WTP at each location per treatment group

Chapter 4 A COMPARISON OF EX ANTE AND EX POST CONSEQUENTIALITY EFFICACY TO REDUCE HYPOTHETICAL BIAS

4.1 Abstract

Ascertaining or inducing policy consequentiality is key for incentive-compatible responses from participants in stated preference approaches. Understanding policy consequentiality has not occurred in conjunction with a treatment of real payment in a field survey, so a true measure of consequentiality's effect on Hypothetical Bias does not exist. We implement ex ante consequentiality, ex post consequentiality, and cheap talk in hypothetical elicitations as well as a real elicitation and compare WTP results across all treatments. We find that the ex ante consequentiality treatment increases WTP relative to both real and hypothetical treatment groups and induces more respondents to select the opt-out alternative less frequently. Conversely, using ex post consequentiality answers to exclude inconsequential respondents was effective at removing differences across treatments in WTP measures. Using ex post consequentiality also generally increases WTP across all treatments, as has been previously observed, including in the real payment treatment. Our results of ex ante consequentiality illustrate that its usefulness to mitigate HB remains uncertain and additional investigation is warranted. Lastly, there was limited evidence of HB. Relatively minimal extent of HB in the hypothetical treatment without additional HB mitigation may explain why some HB reduction treatments such as Cheap Talk have appeared ineffective in past studies since these

studies do not implement a treatment involving real payment to determine whether HB was initially present.

4.2 Introduction

The study of consequentiality has grown in prominence in environmental economics and has begun to be recognized in other related fields, both as an explanation of Hypothetical Bias (HB) and as a mechanism to reduce HB. Herriges, et al. (2010) specify two aspects of consequentiality: 1) the perceived likelihood that the respondent's answer affects the outcome of interest ("policy consequentiality") and 2) the perceived likelihood that the respondent must pay given their answer ("payment consequentiality"). The presence of both constitutes "strong consequentiality." Studies in lab settings can easily incorporate elicitation mechanisms and implementation rules such that respondents know there is a non-zero chance of their answer affecting the policy outcome or chance of payment. In these studies, payment consequentiality is imposed on the respondent.

In many applied valuation studies, payment consequentiality is impractical, where evoking actual payment outside of the lab is extremely difficult such as studies where the good is not readily available or in self-administered online surveys. On the other hand, policy consequentiality may still be useful to reduce HB and ascertain more accurate welfare estimates, most commonly Willingness to Pay (WTP). In an applied setting where imposing payment is not possible or is

impractical, the question remains whether the respondent's belief in consequentiality can affect hypothetical WTP.

One of the major goals of this study is to examine the effectiveness of consequentiality in reducing HB. We further define and compare the effect of both ex ante and ex post consequentiality on respondents' WTP. The comparison is based on a field survey where actual WTP is also obtained in a treatment involving actual payment. This type of comparison has not been previously conducted in the literature.

Ex ante treatments provide survey or experiment respondents to consequentiality treatment prior to the WTP elicitation question in a survey or experiment. Much in the same way that Cheap Talk (CT) scripts are presented to respondents prior to valuation, the use of ex ante Consequentiality scripts has grown as a means of mitigating HB. Before Carson and Groves (2007), a common survey practice was to include a reminder on the survey's potential influence on policy. In fact, a number of studies (e.g. Bosworth, et al. (2015), Donfouet, et al. (2011), Hensher, et al. (2005) and Yao, et al. (2014)) have explicitly considered and included language to evoke policy consequentiality in their pre-valuation scripts to *all* respondents, not just a subset.

The other major method of employing consequentiality in stated preference valuations is an ex post correction, which adjusts the data/respondents analyzed based on responses collected in the survey or experiment after the valuation elicitation. The most common ex post approach asks respondents how likely the results of the study will affect broader policy

decisions. Much in the same way as certainty follow-up studies, the sample analyzed is adjusted based on these ex post consequentiality answers. Respondents' answers indicate whether their responses were given conditional on sufficient perceived consequentiality, otherwise their responses will be excluded in the data. Previous studies have shown that respondents who believe the study has no effect on policy generate WTP different from those who believe there will be some effect on policy (Herriges, et al., 2010, Interis and Petrolia, 2014).

In this study, we examine policy consequentiality in a field survey where a real payment treatment is also adopted. This generates WTP measures that can be used to examine our primary goals:

(1) Compare the effects of ex ante and ex post consequentiality on HB

(2) Compare the effect of Cheap Talk to ex ante consequentiality

As a supplemental goal, we also consider the effects on WTP of sample adjustments from removing inconsequential respondents compared to protest respondents, an avenue of research that has not been previously explored. We study these questions in the context of implementing a real and hypothetical choice experiment (CE) for monarch butterfly conservation in a field survey. Our results can inform future stated preference research design on the potential benefits of including CT, ex ante and/or ex post consequentiality to reduce HB.

4.3 Background

Studies considering the effect of consequentiality largely grew in the wake of Carson and Groves (2007).[43] They formalized the importance of elicitation format as well as respondent beliefs as a means of ensuring incentive-compatible choices. In the following sections, we broadly categorize consequentiality studies as one of three study designs: 1) ex ante binding, 2) ex ante non-binding, and 3) ex post.

Ex ante binding studies explicitly notify respondents that there is a non-zero probability of provision or payment. They include at least one treatment that is not purely hypothetical, in the sense that provision and payment are determined by two factors: the first is that a costly choice was selected, often times by majority of the respondents; and the second is that whether provision and payment are binding is determined by a random mechanism such as toss of a coin. Some studies change the probability of either or both to characterize the effect of consequentiality.

These critical requirements of real payment and provision consequentiality often mean that such studies are usually restricted to controlled lab or field experiments. This vein of consequentiality research is prolific, outlined extensively in Poe and Vossler (2011) as well as Carson, et al. (2014).

[43]Some empirical work existed beforehand such as Bulte et al. (2005) and Cummings and Taylor (1998). They cite earlier versions of Carson and Groves' work, which first became available in the mid 90's.

A number of studies have investigated ex ante binding consequentiality using induced value experiments, where the respondent's value in the experiment is assigned, rather than homegrown. Examples include Burton, et al. (2007), Collins and Vossler (2009), Taylor, et al. (2001), Vossler and McKee (2006), and Polomé (2003). Poe and Vossler (2011) show that for the combined results of the first four studies, 92% of the induced-value votes were "right" based on the voter's expected payoff, and many "wrong" votes explained by those whose expected payoffs were near $0. Conversely, Polomé (2003) found that over 50% of respondents did not truthfully reveal WTP in a referendum.

On the other hand, studies of homegrown values are more similar to typical stated preference approaches. In these cases, respondents must decide their value for themselves before answering the elicitation questions. Mitani and Flores (2014) used 30 different combinations of payment and provision probability, ranging from a purely hypothetical (0% probability of payment and provision) to a purely real (100% chance of payment and provision). They empirically demonstrate that higher probability of payment (provision) decreases (increases) contributions and that respondent risk-attitudes also affect payments. Vossler and Evans (2009) found that referenda with various types of advisory consequentiality treatments produce equivalent results as a real referenda. Landry and List (2007) as well as Barrage and Lee (2010) both used a 50% probability that respondent choices would be binding and in both cases found consequential WTP was equivalent to real WTP.

In addition to comparison of consequential and real treatments, some have investigated effects of varying degrees of consequentiality treatments. Theory predicts that as long as respondents believe there is a non-zero probability of provision *and* payment, even if only trivially greater than 0, then their dominant strategy is to answer truthfully, a so-called knife-edge result. Cummings and Taylor (1998) found that WTP with a 75% probability of being binding was equal to the real treatment, whereas 0% (purely hypothetical), 25% and 50% were still greater. The results of the studies mentioned suggest that relying on a knife-edge remains an open question.

Ex ante, non-binding consequentiality explicitly tells respondents that their individual responses matter before the information and results of the study will be communicated to public officials. As mentioned earlier, such appeals have been included to various degrees in many valuation surveys as a way to increase respondent cooperation. Relative to the ex ante, binding consequentiality papers above, this approach is much more feasible to employ in typical stated preference surveys conducted with or without close interaction between the researcher and the respondents.

We found the fewest papers for ex ante, non-binding consequentiality. An early, prominent example comes from Bulte, et al. (2005) who found that ex ante consequentiality scripts were at least as effective as CT in reducing WTP, based on the following script: *"Note: the results of this study will be made available to policy makers, and could serve as a guide for future decisions with respect to taxation for this purpose. It is important that you think before answering the*

question." More recently, Kemper, et al. (2016) found that an ex ante consequentiality script for a CE of chicken breast significantly lowered WTP for some attributes.[44]

Vossler and Evans (2009) state their implicit advisory treatment is akin to a policy consequentiality treatment because respondents know their vote affects outcomes, but exactly how remains unknown to the respondent, much in the same way as a typical survey.[45] Their implicit advisory treatment was equivalent to the baseline, real payment treatment.

Three unpublished works (Drichoutis, et al., 2015, Hidano, et al., 2005, Williams, 2013) all found that ex ante consequentiality scripts did not affect values for strawberries, climate change and transitioning-into-certified-organic apples, respectively. Lastly, when Lewis et al. (2016) employed a script[46], they

[44] An important caveat is that part of their consequentiality script contains elements of Cheap Talk and was adapted from two previous Cheap Talk scripts (List, 2001; de-Magistris et al., 2013). In our assessment, roughly one-third of the 187-word script focused on policy consequentiality. The remaining two-thirds focused on a budget reminder often employed in Cheap Talk.

[45] One could argue that even if students did not know the explicit decision rule, this treatment is more similar to an ex ante binding treatment since students know that as a university lab-experiment, policy outcomes are enforceable.

[46] The script was three bullet points as follows: "IMPORTANT: •Your responses will be used to assist policy makers in determining genetically modified labeling practices and in determining how much foreign sugar should enter into the United States. •Based on your preferences, policy makers could determine whether foreign sugar should be able to enter into the United States and at what rate. •Your decisions could also help policy makers determine if genetically modified foods should be labeled."

found that respondents were significantly more likely to select a costly option rather than an opt-out alternative, but had no difference in WTP.

Most importantly, it appears that none of the ex ante studies mentioned also included a real valuation group, so it is impossible to establish the extent of HB. Even while the effect of consequentiality scripts to reduce HB is mixed, as mentioned previously, studies are beginning to show them to all respondents as one of a number of safeguards against potential HB.

The last collection of studies use ex post consequentiality to correct WTP, with numerous studies appearing after 2009. Its most common form is a Likert-style question after the value elicitation that asks the respondent the degree to which they believe the results of the survey may affect policy. Ex post consequentiality questions have also been a central method to test for knife-edge results. Applying the knife-edge supposition in context of an ex post consequentiality question implies that those who believe there is a trivial but non-zero chance of their response affecting outcomes should generate truthful, incentive-compatible values equal to those who have a much higher chance that the survey may affect policy. Said differently, the best strategy for respondents is to give truthful answers for any non-zero chance of becoming true.

In a few studies, respondents faced a real value elicitation with an ex post consequential follow-up. Vossler and Watson (2013) compared hypothetical mail respondents to the results of an actual public referendum for conservation and perseveration efforts in a Massachusetts municipality. They find that inconsequential survey respondents underreport their WTP relative to real

results. Broadbent et al. (2010)'s comparison of real payment (determined by a 1 out of 20 draw from a bingo cage) and hypothetical payment groups found that WTP based on consequential respondents had no effect on reducing HB. In both cases, an ex ante consequentiality treatment was not included, so no comparison of HB from ex ante or ex post methods is possible.

Most cases of ex post consequentiality only implement hypothetical valuations. A number of papers find that higher ex post consequentiality lead to higher WTP (Interis and Petrolia, 2014, Li, et al., 2016, Nepal, et al., 2009, Vásquez and Franceschi, 2013, Vásquez, et al., 2009). Hwang, et al. (2014) studied how the respondent's perceived consequentiality affected opt-out rates. They did not consider the effect on WTP, though higher opt-out rates usually correspond to lower WTP. Most recently, Groothuis, et al. (2017) studies both policy and payment consequentiality and found the perceived consequentiality was endogenously determined with the tax amount quoted in a referendum.

Lastly, a few studies implemented both ex ante policy consequentiality and ex post consequentiality, of which we know of three, Herriges, et al. (2010), Drichoutis, et al. (2015), and Lewis, et al. (2016). Herriges et al. used the ex ante consequentiality treatment as a means of controlling endogeneity in the ex post consequential beliefs of respondents. They find that the treatment did influence ex post consequentiality among respondents, but did not include it in their model of WTP. Drichoutis, et al. (2015) found that neither consequentiality nor CT scripts affected WTP estimates. While neither was significant, upon closer inspection, CT had a negative sign (i.e. reduced WTP) across specifications,

whereas ex ante consequentiality seemed to increased it in some circumstances. Similarly in Lewis, et al. (2016), both ex ante and ex post consequentiality had little effect on WTP for both sugar and soft drinks. In all three cases, a real payment treatment was not included, so a true measure of HB could not be established.

4.4 Research Design and Data Collection

To study the effect of ex ante and ex post consequentiality on WTP, we conducted a CE in a field survey on the valuation of monarch butterflies. This valuation establishes HB by implementing both hypothetical and real payment treatments. Among respondents assigned to the hypothetical payment treatment, they either received an ex ante consequentiality script or a CT script. All respondents answer the ex post consequentiality questions.

Our consequentiality script is as follows: *"Please note that state and local administrators and policymakers are aware of this study and anticipate using its results to serve as a guide for decisions related to butterfly and pollinator conservation efforts in the near future throughout Kentucky. Your answers can affect the policymakers' priorities and decisions. So in the next six situations that you will see shortly, carefully consider each option and make your preferred choice. Remember, your opinion counts."*

Our script is approximately 72 words, comparable to the 70 and 40 word script of Lewis, et al. (2016) and Bulte, et al. (2005), respectively. It also follows their strategy of employing a nonspecific appeal to respondents on their choices

being communicated to policymakers and guiding future decisions. This was done to remain general enough to apply to any number of valuation studies.

To be more comparable to the consequentiality script, the CT script was of similar length, per the following: *"For hypothetical questions like these, people often say they are willing to donate more for conservation than they would actually pay in a real donation using their own money. People may not consider the money they are giving up since it's easy to be generous when a real payment isn't being made. Even though your choices in the six situations are hypothetical, please imagine that if one of the situations were randomly selected, that you would actually donate the amount based on the option you've chosen in that situation."*

As defined by Ami, et al. (2011), the script is 'positive' in that it specifically indicates that people tend to overstate their values in hypothetical elicitations, and 'light,' which means that no quantitative information is given. Our script is based on the short scripts previously implemented by Aadland and Caplan (2006) and Carlsson, et al. (2005).

For our ex post consequentiality question, we asked: "How likely do you think it is that the results of this survey will shape the direction of future public policy of butterfly conservation in Kentucky?" Respondents could answer "Very Likely," "Likely," "Very Unlikely," "Unlikely," or "I don't Know". We define someone as being policy consequential if they select "Very Likely" or "Likely".

While previous works have extensively considered respondent beliefs in the credibility of the survey to affect policy, we believe that a separate but related

issue is whether they believe in the credibility of the good itself, especially public goods. It is possible that even after a lengthy explanation; some may not believe the described mechanism can be effective, which we label product consequentiality. To measure product consequentiality, we also asked whether "installing butterfly plants can actually help butterfly conservation?" . In other words, do respondents believe in the credibility of the good itself, which is related to but distinct from policy consequentiality. We use a similar definition for product consequentiality as in policy consequentiality.

This experimental design was in the context of CE to value conservation support for monarch butterflies (*Danaus plexippus*). Monarchs are among the most well-known butterflies in the United States, known for their vibrant orange color pattern. In the past two decades, the monarch population has plummeted to a fraction of its former size (Brower, et al., 2012, Jepsen, et al., 2015). A petition to list the Monarch butterfly as an endangered species was received in December 2014, and a final decision is due in June 2019. The CE focused on accepting donations to provide additional plants and habitat to support monarch butterflies. This stems from evidence that the monarch's decline is due to more limited resources (Flockhart, et al., 2015, Pleasants and Oberhauser, 2013)[47] and because it is a widely-accepted mechanism to support monarch conservation. Additionally, installing plants has the benefit of being tangible and divisible, making the good quasi-public, mitigating some of the free-riding issues of a typical public good.

[47] Inamine et al. (2016) demonstrate that this issue is not settled among entomologists.

Participants were told that all donations go towards the purchase and installation of plant seedlings, each at a cost of $1, based on local estimates for purchasing seedlings in bulk. The CE attributes themselves focused on the location, accessibility, and waystation designation of the restoration sites, with each listed in **Table 4-1**.

The CE is based on a full factorial design, using 36 two-alternative choice sets, blocked into six groups of six choice sets. Each respondent participated in one of six groups of choice sets. Upon completing the six choice sets, respondents in the real payment treatment group rolled a 6-sided die to determine which choice set would be binding and, if appropriate, made their donation immediately after the roll. An example choice set appears in **Figure 4-1**. To improve comprehension, CE instructions were conveyed as a short video with an example choice task.

The CE included a follow-up question queried those who choose not to donate in all six situations to identify protest respondents based on Diffendorfer, et al. (2014). Those who did not feel it was their responsibility to protect butterflies, who did not think the program would be effective, or who did not trust the money would be spent on butterfly conservation were labelled protesters and their WTP is categorized as protest zeroes.

Surveys were collected during the summer of 2016 at almost three dozen unique locations on 51 separate occasions in and around Lexington, KY. Each day of the week and each time of day was surveyed multiple times, but responses tend to come from weekday afternoons/evenings and weekends.

These collections at occurred at a variety of locations and events such as a county fair, a movie in the park, at playgrounds, at sports events, and at music festivals. This makes it possible for the sample to be qualitatively similar to the general Lexington population, though we do not claim it is representative of the broader US population.

After a potential respondent agreed to participate, they began by completing a separate exercise to earn $10. This is to allow the respondents to treat the money as earned instead of windfall/house money, the latter of which may distort WTP (Clark, 2002, Loureiro, et al., 2003). Both hypothetical and real payment respondents received five $1 bills and one $5 bill. Respondents in the real payment treatment were told at the beginning the survey that they would have a chance, but were not obligated, to make a real donation during the survey. All respondents then proceeded to complete the survey. One survey administrator was present at all events as well as a small group of rotating assistants, reducing potential interviewer bias.

4.5 Econometric Approach

Discrete choice models are based upon Random Utility Theory, which describes a person's utility from a particular good being composed of observable and unobservable components to the analyst (McFadden, 1973). Equation 1 shows that individual *i* derives utility from selecting alternative *j* in choice set *t* with observable attributes X_{ijt}, payment variable c_{ijt}, an unobservable

component $\boldsymbol{\varepsilon}$, as well as a scale parameter *k* associated with the unobservable component:

$$\boldsymbol{U_{ijt} = -(\alpha_i/k_i)\, c_{ijt} + (\beta_i/k_i)' X_{ijt} + \varepsilon_{ijt}} \quad (1)$$

Among the coefficients to be estimated, *a* represents the effect of change in price while the vector for β yields the estimated effect of various attributes on their choice, and indexed by *i* to showcase that the effect of attributes varies across individuals, one of the primary advantages of conducting a mixed logit model based on (1).

The above specifications represent a model in parameter space. The scale parameter, k_i, is inherent to but not separately identifiable in the model, and is assumed to be fixed, such that the unobservable component's variance is equal across respondents (i.e. homoskedasticity). This issue of scale has two important implications: 1) comparing coefficient estimates across samples is inappropriate due to scale differences, and 2) that the variability in unobserved utility is the same for all respondents, which can potentially bias other coefficient estimates in the model. A number of model extensions exist to relax various assumptions such as modelling choices in WTP-space or using generalized multinomial logit. Given the limited sample size, model convergence in these more general models was infeasible. As such, our models are based on a standard mixed logit model with a fixed price coefficient and normally distributed non-price coefficients. WTP is calculated by multiplying negative one by the ratio

of the attribute coefficient to the price coefficient. WTP inherently removes the scaling factor, making comparison appropriate across samples.

Equation 2 reflects the model specification in our application for each respondent. To facilitate exposition, subscript *i* is omitted. Parameters β_l reflecting the taste for each respective attribute in the reference group composed of respondents in the treatment group that entails real payment. Dummy variables $Treatment_m$ indicates the *m* treatment groups and δ_m representing the change in preferences for these *m* hypothetical treatment groups (Control, Ex Ante Consequential, and CT). These interactions provide the primary mechanism for testing differences across treatments.

$$\begin{aligned} \boldsymbol{U_{jt} = \alpha Donation\ Amount_{jt} + \beta_1 OptOut_{jt} + \beta_2 Elizabethtown_{jt}} \\ \boldsymbol{+ \beta_3 Lexington_{jt} + \beta_4 Public_{jt} + \beta_5 Waystation_{jt}} \\ \boldsymbol{+ \sum_m \delta_{m1}(Donation\ Amount * Treatment_m) + \sum_m \delta_{m2}(Optout *} \\ \boldsymbol{Treatment_m) + \sum_m \delta_{m3}(Elizabethtown * Treatment_m) +} \\ \boldsymbol{\sum_m \delta_{m4}(Lexington * Treatment_m) + \sum_m \delta_{m5}(Public * Treatment_m) +} \\ \boldsymbol{\sum_m \delta_{m6}(Waystation * Treatment_m) + e_{jt}} \end{aligned} \quad (2)$$

The opt-out constant represents a choice not to donate in a particular situation. It usually represents the disutility of being unable to consume the offered good with the base level of the various attributes not explicitly captured by other variables in the utility function. In our case, this is the installation of plants in Paducah, KY in a private location without the waystation designation (presumably the least valuable alternative possible).

We use 500 Halton draws in parameter space. Based on log-likelihood, the mixed logit models are superior to conditional logit models (not presented). Lastly, we rely on the delta method for a number of post-estimation comparisons of model results.

4.6 Results

4.6.1 Sample Description and Summary Statistics

The field survey yielded 397 useable responses. **Table 4-2** provides summary statistics of various socioeconomic factors across the four treatments. Comparing to information on the relevant metropolitan statistical area, we observe that population information and the combined treatments are largely similar. It does contain some differences, such as attaining more education relative to the general population and more frequently having minors in the household. In examining each of the four treatments, there are no statistically significant socioeconomic differences, demonstrating that randomization of the treatments was successfully implemented.

Table 4-3 showcases respondent answers per treatment in terms of the proportion who believed in policy and product consequentiality, the proportion who protested, and the proportion of opt-outs in the CE. With respect to policy consequentiality, roughly two-thirds of respondents were policy consequential, with no significant differences across treatments.[48] We find that about nine-tenths

[48] A potential shortcoming of CT scripts is that by pointing out the survey is hypothetical, CT scripts could *reduce* the ex post consequentiality of respondents. The percentage

of respondents believed that installing plants could help butterfly conservation, lending credence to the product consequentiality.

With respect to the frequency of opting-out and protest respondents, there are a number of significant differences. Respondents in the real payment treatment group are significantly more likely to opt-out compared to all three hypothetical treatments, an indication of HB. While the opt-out rate is quite high for real, it had significantly fewer protest respondents versus hypothetical and ex ante consequential while CT had comparable rates of protestors.

Table 4-4 summarizes the impact of the different treatments. As a means of comparison, we use the answers in the protest question and ex post policy consequentiality question to generate alternative model results based on their exclusion. Those who did not feel it was their responsibility to protect butterflies, who did not think the program would be effective, or who did not trust the money would be spent on butterfly conservation are excluded in the No Protest model results (Model II). Those who answer “Very Unlikely” “Unlikely” or “I don’t know” to the ex post consequentiality question are excluded in the Consequential Only (III) model results.

For the final model results (Model IV), the sample excludes those who are jointly protestors *and* inconsequential. Whereas a more conservative practice is to drop *either* protestors or inconsequential respondents, as in Model II and III, respectively, this can potentially represent a substantial proportion of the sample.

69.0% of ex post consequential respondents in CT is slightly, but not significantly higher than real-payment (67.3%) or hypothetical (65.5%) treatments.

In environmental applications set in the field, such reductions can be detrimental to model efficiency. Model IV shows how much HB is affected by excluding only the worst offenders, those who protest and as the same time do not believe in the policy consequentiality of the study.

4.6.2 Model Results

To begin with, we examine the non-interacted variables in models I to IV. These coefficients represent those under the real payment treatment group. We observe that higher donation requests significantly reduce the likelihood of the costly alternative being chosen and a positive coefficient for the opt-out alternative. With respect to the location of the sites for the potential plants we observe that relative to the baseline (Paducah, furthest away), respondents are indifferent to adding plants in Elizabethtown, and significantly prefer to add plants in Lexington. Given past evidence of distance decay (León, et al., 2016, Sutherland and Walsh, 1985), this pattern is unsurprising.

The accessibility was also significant in every model. This means that respondents value the option value of visiting butterfly conservations sites. The monarch waystation certification was unimportant to respondents. This may be due to the fact that certification can only take place after the plants have been installed, which means there is no additional benefit to attracting additional butterfly conservation prior to becoming certified. Finally, based on a normal distribution assumption, we see that there is significant heterogeneity in all attributes in respondent preferences.

Most importantly, the direction and significance of these attributes is robust across the four models. To test the impact of different HB reduction treatments, we originally included interactions between all attributes and all treatment effects. However, for attributes Elizabethtown and Waystation, their respective interactions with all treatments effects were insignificant in all models. Given our limited sample size, efficiency is crucial, so we excluded these interactions from all future analysis.

We now turn our attention to the interactions of the three hypothetical treatments specifically in Model I, labelled Hypo, Conseq, and CTalk, followed by the variable name, in **Table 4-4**. Relatively few interactions are significant. What we do observe is that the price interaction for hypothetical and consequential treatments is significant and positive, which means that these groups are less price sensitive than those in the real payment treatment. It also inherently means that WTP for these treatments' corresponding attributes is significantly higher than in the real payment treatment, which is evidence of HB. The price interaction for CT was not significant, indicating that price sensitivity are similar in the CT and real payment treatments.

Specifically with respect to the opt-out, recall that respondents in the real payment treatment generally favor avoiding a contribution to butterfly conservation described in the baseline case.[49] The opt-out interaction for

[49] This is not saying that people prefer that butterflies not exist. Instead, because the donation mechanism is for plants that support butterfly conservation, it is indicative of the willingness to purchase the plants in the conditions specified in the least desirable baseline case.

hypothetical, ex ante consequential, and CT are all negative, indicating respondents in the hypothetical payment treatments are relatively less inclined to opt-out, though this is only significant in the hypothetical payment treatment and the hypothetical with consequentiality treatment. Furthermore, combining the opt-out constant with the opt-out interactions for the three hypothetical treatments still produces positive opt-out coefficients, but not significantly different from 0 (Hypothetical: 2.369-2.194=.175, p=.869; Consequential: 2.369-1.495=.874, p=.379; CT: 2.369-1.569=.800, p=.413). This smaller likelihood of opting out in hypothetical and ex ante consequential treatments can be interpreted as support for the result in **Table 4-4**, which is that respondents in a real payment treatment tend to choose the opt-out more often than in hypothetical treatments. In model I, significant interactions for price or the attributes indicate HB on the intensive margin (Ladenburg and Olsen, 2014, Meyerhoff and Liebe, 2009).

Next, we examine models II, III, and IV. We see that the model fit of all four models is generally quite similar, based on the per unit log-likelihood used in each model. With respect to removing protest respondents (II: No Protests) we see that by removing protestors, some HB in the hypothetical treatment is ameliorated because the interaction for the donation amount is no longer significant, removing the implicit HB in the WTP for the attributes. On the other hand, there is still HB shown through the significant opt-out interaction. Because the donation and opt-out interaction remain significant in the ex ante consequential treatment, HB appears not to be mitigated at all. Lastly, CT still has no significant interactions, making the CT treatment statistically equivalent to

real payment treatment. In the sample featuring ex post consequential respondents only (Model III), none of the treatment interactions for hypothetical, ex ante consequential, or CT are significant. This means that by removing inconsequential respondents, we mitigate HB, both with respect to WTP in the attributes as well as in the rates of opt-outing from donating.

Before inspecting the results of model IV, recall that about a ninth of respondents can be excluded due to protesting and another third of respondents could be excluded due to being inconsequential respondents. In fact, a union of the two groups represents approximately 39% of the sample. Excluding such a large proportion of the data negatively affects statistical efficiency. This motivates the usefulness of model IV, which excludes only those who both protested and believed the study as inconsequential, which represents about 5% of the sample, a more palatable set to exclude. Even with this minimal exclusion rule, we see that the interactions for the hypothetical treatment are not significant, albeit quite close to a p-value=.1. On the other hand, it still appears that the ex ante consequentiality script tends to exacerbate HB, based on its significant price interaction. There is also some evidence of HB for CT since Public accessibility is significant. On the other hand, the extent of HB in CT may be considered less severe versus ex ante consequentiality since the former only exhibits HB in a single attribute, while the significance of the price interaction in the latter means there is HB in all attributes.

In summary, based on our sample, it appears that providing an ex ante consequentiality script can increase attribute WTP, counter to the results Bulte,

et al. (2005), as well as in market participation, which is similar to Lewis, et al. (2016). Further, removing inconsequential respondents can serve to mitigate HB the most of the ex post exclusion strategies. Because ex ante consequentiality tends to increase WTP, it inherently means it performs poorly relative to ex post consequentiality as well as to CT, answering two of our primary objectives. Lastly, it is worth mentioning that interactions for the hypothetical treatment group fluctuate in their statistical significance. This means that the evidence of HB is not particularly acute. Since there is modest HB, this can explain why CT is not particularly effective in ours. We also argue that this could be one of the reasons why CT and some other HB-alleviating methods are not found to be effective in some previous research, as in these studies a real payment treatment was often not implemented. The finding that a HB-alleviating treatment is ineffective or even counter-effective is only established based on the judgment whether the WTP measures generated in these HB-alleviating treatments are less than what they are in the purely hypothetical treatment. Our results present one case that when in reality HB does not persist, HB-reducing treatments may not function as expected.

As a second opportunity for comparison, we consider the WTP in each of the four models, as in **Table 4-5**. WTP removes the scale factor inherent to each model such that direct comparison across the four models is inappropriate. The WTP estimates and p-values are all based on the delta method. For example, the significant difference between hypothetical public and ex ante consequential-public in model I is based on the WTP for public in hypothetical ($1.70) and

consequential ($6.48) by combining the appropriate interaction terms. In this case, WTP is -1*(1.734-.757)/(0.844+0.269)=$1.70 compared to -1*((1.734+1.044)/(-.844+0.415)=$6.48 in the two respective groups.

In order to examine the effect of ex ante consequentiality, we consider changes in WTP within each model. To begin with, differences between WTP in the hypothetical and real payment treatment groups constitutes the traditional measure of HB. In model I, there are no significant differences between hypothetical and real payment treatment groups, meaning there is little evidence of HB with respect to WTP. In model II and III, there is some evidence of HB, but only in the opt-out alternative. Conversely, every model shows evidence of a significant difference in real payment and ex ante consequential, and three of the four models show a significant difference between real payment and CT.

This suggests only modest evidence of HB in the hypothetical treatment group, but that introducing ex ante consequential scripts and CT may in fact worsen the extent of HB. This point is further supported because among the hypothetical treatments, WTP tends to be lowest in the control group where no HB-reducing treatments were used. In two cases, ex ante consequentiality WTP is significantly higher than hypothetical WTP. This is similar to the model results from **Table 4-4** that show that ex ante consequentiality may exacerbate HB. A similar trend exists for CT, though without significance.

As part of our second goal, we more closely examine ex ante consequentiality versus CT. While there is evidence both treatments have higher WTP relative to real payment and to the baseline hypothetical group, there is no

such within-model evidence that the two are different from each other based on WTP.

To examine the effect of ex post consequentiality, we must examine WTP across the four models. Relative to WTP for all respondents (Model I), WTP tends to increase in the no-protest and consequential only results. This is expected since removing protestors correspondingly means removing choice sets where no donation was made. Similarly, previous work has demonstrated greater WTP among consequential respondents (Herriges, et al., 2010, Interis and Petrolia, 2014). WTP in model IV again tends to be higher than Model I, but quite similar to the results of Model II and III. Lastly, there is no clear pattern to suggest that removing protest respondents versus inconsequential respondents is better at mitigating HB in WTP.[50]

4.7 Discussion

Using a field survey on monarch butterfly conservation, we conduct a split-sample experiment to examine ex ante and ex post consequentiality strategies to mitigate HB, including a real elicitation, which has previously not been included before in studies on policy consequentiality. To begin, respondents are

[50] This is further supported by a series of comparisons of WTP across the four model specifications using the Poe combinatorial test (Poe et al., 2005) in conjunction with the Krinsky-Robb procedure. Given 14 WTP values generated per model and 4 models, there are a total of 84 pairwise comparisons (Krinsky and Robb, 1986). None of the tests had marginal significance and only four tests yielded a p-value<.2.

generally unwilling to donate to support monarch conservation if the location of the effort and accessibility are not favorable, but they do show significant WTP if restoration occurs nearby or if the location is publicly accessible.

In comparing the various hypothetical treatment groups to the treatment with real payment, we find there is evidence of HB, but among the three treatments involving hypothetical payments, the control group has some evidence of HB in the model results, but relatively little in WTP. Conversely, CT tends to have little evidence of HB in the model results, but much more so in the WTP results. The fact that when HB is not persistent, HB-reducing treatments, such as CT, may not function as expected may serve as some evidence as to why CT is found to be ineffectual in some previous studies where no treatments with real payment were implemented.

For ex ante-consequentiality, there is evidence of HB. We find some support to suggest that ex ante consequentiality increases the WTP for certain attributes. While this appears to counter to Bulte, et al. (2005), most ex ante consequentiality studies have found it had no effect. More intuitively, the purpose of ex ante consequentiality scripts is to increase participants' belief in the consequentiality of the survey and study. Others have found that those who are ex post consequential tend to have higher WTP. By extension, it seems reasonable to expect that ex ante consequentiality increases WTP or market participation.

We observe that ex ante consequentiality leads to the opt-out choice to be selected much less frequently, a phenomenon previously observed by Lewis, et

al. (2016), indicating the ex ante consequentiality may influence market participation.

This means that both ex ante and ex post consequentiality methods tend to increase WTP. For ex ante consequentiality, it is an increase WTP relative to real payment treatment and the hypothetical treatment. For ex post consequentiality, higher WTP is relative to a model of all respondents. Lastly, ex ante consequentiality and cheap talk are not significantly different from each other in terms of their interactions in the underlying model results or in WTP.

Table 4-1 Choice Experiment Attributes and Levels (Same as Table 3-1)

Attribute	Description	Levels
Location	Potential sites in Kentucky to install butterfly plants	1. Lexington (Fayette County KY) 2. Elizabethtown (Hardin County KY): 85 miles away from Lexington (1.5 hour drive) 3. Paducah* (McCracken County KY): 350 miles away from Lexington (4 hour drive)
Accessibility	Public's ability to visit site	1. Open: habitat accessible and viewable by the public, such as public parks 2. Closed*: habitat inaccessible nor viewable by the public, such as a private farmland
Waystation	Inclusion in national waystation program	1. Certified: Waystation is certified and Waystation Sign is installed. 2. Not Certified*: Habitat is not a certified Waystation nor is a Waystation Sign installed.
Donation	Amount of money to support butterfly plants	$1, $5, $10

* indicates reference category in CE

Table 4-2 Sample Summary Statistics

Variable	Population[1]	Total Sample	Hypo	H-Conseq	H-Cheap Talk	Real
N						
Age[2]	36.0	38.4	38.5	39.0	35.8	39.2
18-24	18.6	17.4	19.5	13.8	23.9	15.2
25-34	19.4	31.1	28.7	28.8	36.6	31.0
35-44	16.8	20.0	20.7	23.8	12.7	20.9
45-54	16.0	15.7	12.6	21.3	12.7	15.8
55-64	14.6	12.1	13.8	10.0	12.7	12.0
65+	14.7	3.8	4.6	2.5	1.4	5.1
Chi-2(15)=11.8, p =.69[3]						
Male	48.6	44.1	37.9	45.0	39.4	49.1
Female	51.4	55.9	62.1	55.0	60.6	50.9
Chi-2(3)= 3.6, p=.31						
Education						
High school or less	30.0	21.5	19.8	17.5	25.7	22.6
Some college	27.4	24.1	30.2	20.0	21.4	23.9
Bachelor's degree	23.6	27.9	29.1	35.0	20.0	27.0
Graduate/professional	17.0	26.6	20.9	27.5	32.9	26.4
Chi-2(9)=8.7, p=.47						
White	75.6	80.7	83.8	84.3	77.9	78.4
Black/African American	14.4	16.2	15.0	11.4	20.6	17.3
Asian	3.6	3.1	1.3	4.3	1.6	4.3
Chi-2(6)=4.6, p=.60						
Minors at home	28.9	47.8	47.1	55.0	45.1	44.9
Chi-2(3)=2.4, p=.50						
Single, never married	38.8	33.3	37.9	25.0	43.7	30.2
Married	41.1	50.9	50.8	58.8	39.4	52.2
Other marital status	20.1	15.9	11.5	16.3	16.9	17.6
Chi-2(6)=9.24, p=.16						
Median Income[4]	$47,968	$42,500	$42,500	$62,500	$42,500	$42,500

[1] Based on U.S. Census Bureau's 2015 American Community Survey 1-Year Estimates of Lexington-Fayette Metropolitan Statistical Area

[2] Age percentages based on population 18 or older.

[3] Chi-square tests examine whether there are significant differences among the four treatments for each group of variables (e.g. age, education, etc.).

[4] Based on mid-point of response

Table 4-3 Proportion of Respondents based on Ex Post Consequentiality and Protests

	Total	Real	H-Control	H-Ex Ante Consequential	H-Cheap Talk	Significant Difference[1]
Number of Respondents	497	159	87	80	71	
% Ex Post Policy Consequential	67.3%	67.3%	65.5%	67.5%	69.0%	
% Ex Post Product Consequential	91.9%	95.0%	87.4%	90.0%	93.0%	A
% of Opt-Outs in Choice Sets	62.8%	70.9%	58.6%	56.7%	56.6%	A,B,C
% of Protest Respondents	11.6%	7.5%	14.9%	20.0%	7.0%	A,B,F

[1] Using a difference in proportions t-test, A, B, C, D, E, and F indicate a significant difference ($p<.1$) between A: Real v. Hypothetical, B: Real v. H-Ex Ante Consequentiality, C: Real v. H-Cheap Talk, D:Hypothetical v. H-Ex Ante Consequentiality, E: Hypothetical v. H-Cheap Talk, F: H-Ex Ante Consequentiality v. H-Cheap Talk, respectively.

Table 4-4 Mixed Logit Model Results

	I: All		II: No Protests		III: Ex Post Consequential Only		IV: Joint No Protests and Consequential Only	
Donation	-0.844***	(0.225)	-0.658***	(0.124)	-0.799***	(0.234)	-0.755***	(0.153)
Opt-Out	2.369***	(0.865)	1.054*	(0.629)	1.958*	(1.15)	1.811**	(0.775)
Elizabethtown	-0.075	(0.408)	-0.102	(0.343)	-0.023	(0.515)	-0.162	(0.397)
Lexington	2.297***	(0.764)	2.056***	(0.634)	2.046*	(1.109)	2.344***	(0.775)
Public	1.734***	(0.627)	1.546***	(0.467)	1.845**	(0.771)	1.557***	(0.51)
Waystation	0.002	(0.347)	0.163	(0.267)	0.202	(0.416)	0.124	(0.31)
Hypo-Donation	0.269*	(0.162)	0.198	(0.121)	0.165	(0.174)	0.228	(0.138)
Hypo-Optout	-2.194*	(1.311)	-2.167**	(1.062)	-2.947	(1.863)	-1.706	(1.096)
Hypo-Lexington	0.349	(1.183)	0.453	(0.977)	0.445	(1.464)	0.242	(1.143)
Hypo-Public	-0.757	(0.93)	-0.551	(0.746)	-0.360	(1.040)	-0.426	(0.832)
Conseq-Donation	0.415**	(0.179)	0.298**	(0.126)	0.300	(0.183)	0.333**	(0.147)
Conseq-Optout	-1.495	(1.256)	-1.738*	(1.011)	-1.927	(1.635)	-1.239	(1.182)
Conseq-Lexington	-0.127	(1.174)	0.447	(1.019)	1.035	(1.503)	-0.062	(1.195)
Conseq-Public	1.044	(0.963)	1.093	(0.798)	1.619	(1.232)	1.041	(0.926)
CTalk-Donation	0.181	(0.159)	0.115	(0.119)	0.242	(0.192)	0.078	(0.139)
CTalk-Optout	-1.569	(1.171)	-0.815	(0.992)	-0.669	(1.547)	-1.822	(1.14)
CTalk-Lexington	1.443	(1.283)	1.354	(1.04)	2.535	(1.672)	1.962	(1.376)
CTalk-Public	1.295	(0.949)	1.037	(0.788)	0.754	(1.135)	1.779*	(1.065)

Table 4-4 Continued Mixed Logit Model Results

	I: All		II: No Protests		III: Ex Post Consequential Only		IV: Joint No Protests and Consequential Only	
Std. Dev.								
Opt-Out	6.353***	(1.44)	4.676***	(0.776)	6.543***	(1.664)	5.869***	(1.107)
Elizabethtown	3.611***	(1.015)	3.168***	(0.725)	4.287***	(1.355)	3.915***	(0.84)
Lexington	5.853***	(1.533)	4.462***	(0.844)	5.083***	(1.361)	5.216***	(1.179)
Public	4.182***	(1.223)	3.163***	(0.722)	4.298***	(1.367)	3.822***	(0.869)
Waystation	-.215***	(0.800)	0.122	(0.877)	2.161*	(1.149)	-0.876	(0.608)
N Choice sets	2382		2106		1602		2256	
LL	-1079.66		-1013.69		-727.84		-1048.97	
LL per choice set	-0.453		-0.481		-0.454		-0.465	

Table 4-5 Willingness to Pay Results based on Table 4-4

	Real WTP	Hypo WTP	Conseq WTP	CTalk WTP	Significant Differences[1]
I: All					
Opt-Out	**2.81**	0.30	2.04	1.21	
Elizabethtown	-0.09				
Lexington	**2.72**	**4.60**	**5.06**	**5.64**	
Public	**2.06**	1.70	**6.48**	**4.57**	B, C, D
Waystation	0.002				
II: No Protests					
Opt-Out	1.60	-2.42	-1.90	0.44	A
Elizabethtown	-0.15				
Lexington	**3.13**	**5.46**	**6.96**	**6.28**	
Public	**2.35**	2.17	**7.34**	**4.58**	B,D
Waystation	0.25				
III: Consequential Only					
Opt-Out	2.45	-1.56	0.06	2.32	A
Elizabethtown	-0.03				
Lexington	**2.56**	**3.93**	**6.18**	**8.23**	C
Public	**2.31**	**2.34**	**6.95**	**4.67**	B
Waystation	0.25				
IV: Joint No Protests and Consequential Only					
Opt-Out	**2.40**	0.20	1.35	-.02	
Elizabethtown	-0.21				
Lexington	**3.10**	**4.90**	**5.41**	**6.36**	C
Public	**2.06**	**2.14**	**6.16**	**4.92**	B,C
Waystation	.16				

Note: All tests are based on the delta method. WTP in **bold** indicate that the value is significantly different from 0 for p<.1.

[1] A, B, C, D, E, and F indicate a significant difference (p<.1) between A: Real v. Hypothetical, B: Real v. H- Ex Ante Consequentiality, C: Real v. H-Cheap Talk, D: Hypothetical v. H-Ex Ante Consequentiality, E: Hypothetical v. H-Cheap Talk, F: H- Ex Ante Consequentiality and v. H-Cheap Talk, respectively.

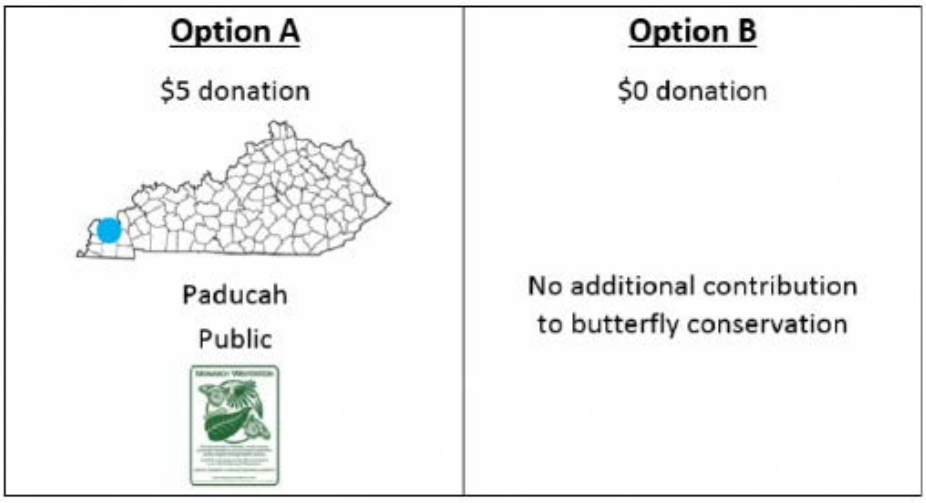

Figure 4-1 Example Choice Set (Same as Figure 3-1)

Chapter 5 Conclusion

5.1 Summary

This dissertation sought to investigate stated preference methods with primary focus on understanding and mitigating Hypothetical Bias (HB). Essay 1 tackles it from the perspective of a meta-analysis; Essay 2 considers unique situations of HB with respect to distance decay and charismatic species; and Essay 3 evaluates the efficacy of consequentiality to reduce HB. Each set of results is discussed below, along with broader implications, and their connection to each other.

With our enhanced dataset on HB both in terms of the number of studies and the characteristics considered, we investigate results of previous meta-analyses as well as expand to new potential factors. We find that the average Calibration Factor (CF) in the trimmed sample is about two, corresponding to the rule of thumb cited by others. On the other hand, about half of all CFs were between .81 and 1.60, while 12.6% of studies had a CF between 3 and 15. This means that a divide-by-two rule cited by some (List and Shogren, 1998, Loomis, 2011) would heavily undervalue the former group, but would still create substantive HB for the latter group.

Researchers should be most concerned with HB when they use auction-type or dichotomous choice elicitations or to value public goods. On the other hand, HB seems unaffected by the survey mode used, endowments, or the use of students.

Another important result of the meta-analysis is the strong evidence that CT, certainty follow-up, and consequentiality reduce HB. In our ancillary ex ante vs ex post model, we see that both are significant in reducing CF, which aligns with Whitehead and Cherry's (2007) suggestion that ex ante and ex post methods could complement each other in reducing HB. Some HB mitigation methods are continuing to grow in prominence, specifically the oath, honest priming, and again, consequentiality. The last category was only included in a limited capacity, but the number of studies focusing on it appears to be increasing[51], so it may be useful to revisit this analysis in the near future.

The results are both promising in that relatively few observations produce extreme CFs, but it is clear opportunities to understand and reduce HB are still necessary. The results of essay 2 and 3 focus on such endeavors.

For essay 2's field survey on monarch and viceroy butterfly conservation, we find a number of results. The purpose of this study was to identify the extent of HB in distance decay as well as whether additional HB was attributable to the charisma of the monarch butterfly. In general, respondents are unwilling to donate to support butterfly conservation. We find that there is distance decay in WTP for both monarch and viceroy conservation, meaning that people prefer to support conservation in their own community compared to a more distant one.

With respect to HB across locations, we observe HB in distance decay for monarchs, but no such HB in distance decay for viceroys. There is still some

[51] 95 articles that included the term 'consequentiality' cited Carson and Groves (2007) through 2014 and there have been 92 such articles from 2015 through spring 2017.

evidence of HB for viceroy though because hypothetical WTP is greater than real WTP in Lexington.

Third, we find that WTP for monarchs exceeds viceroys in the hypothetical treatment, but not in the real treatment; this result suggests there is a hypothetical charisma effect. On the other hand, WTP is equal to support monarchs and viceroys in the real treatments, indicating that the two species are valued equally and no evidence of a real charisma effect. Because the two butterflies are nearly visually identical, this means there is additional HB for monarchs compared to viceroys, evidence that the additional HB is due to a charisma effect. Using our results from the butterfly comparison, we can correct the original annual value of about $410 (in 1995 dollars) for the 16 species in Loomis and White (1996) to a much lower value of about $69. In 2016, this correction is from $644 to about $109.

Because of our definition, we cannot be certain whether the additional HB stems from the rarity of the monarch butterfly or its popularity. Additional investigation can further delineate whether charisma is its own unique explanation of HB in appropriate cases of species valuation, or whether it is a special case of other effects such as social desirability bias.

In order to understand the efficacy of consequentiality, the analysis in essay 3 also relies on the field survey used in essay 2. In it, we find that ex ante consequentiality tends to increase WTP for the attributes as well as increase market participation. It is no different than CT in terms of reducing HB. We also observe that ex post consequentiality tends to increase WTP for attributes across

all treatments, but reduces the extent of HB between the real payment treatment and the three hypothetical treatments.

In comparing the outcomes across the three studies, recall that about half of the observations without HB mitigation yielded values in the hypothetical and real that were similar. Essay 3's results support this in that HB is not always evident in the hypothetical treatment group relative to the real payment group. On the other hand, some of the WTP values from ex ante consequentiality and CT, which tended to exceed the regular hypothetical treatment, were three times larger than their real counterparts. Essay 2 showcases specific instances of HB previously unexplored in any of the studies documented in the meta-analysis. It demonstrates that as more work is done, there are additional characteristics to consider in a future updated meta-analysis.

Broadly, this dissertation demonstrates that HB continues to be an issue and adds to our understanding of its pervasiveness both through meta-analysis and with respect to charismatic species conservation. While researchers should be concerned, they may also take ease in that there is evidence from both the meta-analysis and the field survey that HB may not always be a major concern in the first place. This can explain some of the circumstances when HB mitigation strategies such as CT may appear to 'fail' to reduce HB.

The dissertation illustrates that opportunities to improve CV and stated preference methods remain. Pursuing these refinements is pertinent because while human choices and policies that have already taken place are arguably better measures of "actuality", often times these activities are difficult to measure

using readily-observable or revealed-preference data. This is especially important to non-use values, where stated preference methods targeted at measuring “what if” scenarios are most well-suited for. By improving such techniques, stated preference methods may yet be able to gain wider credibility among economists and the public.

References

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

Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26:271-292.

Ajzen, I., T.C. Brown, and F. Carvajal. 2004. "Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation." *Personality and Social Psychology Bulletin* 30:1108-1121.

Alfnes, F., A.G. Guttormsen, G. Steine, and K. Kolstad. 2006. "Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives." *American Journal of Agricultural Economics* 88:1050-1061.

Alfnes, F., C. Yue, and H.H. Jensen. 2010. "Cognitive dissonance as a means of reducing hypothetical bias." *European Review of Agricultural Economics* 37:147-163.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman. 2008. "Does context matter more for hypothetical than for actual contributions? Evidence from a natural field experiment." *Experimental Economics* 11:299-314.

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* 50:111-130.

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman. 1993. *Report of the NOAA panel on contingent valuation.* Washington, DC: Federal Register.

Balistreri, E., G. McClelland, G. Poe, and W. Schulze. 2001. "Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values." *Environmental and Resource Economics* 18:275-292.

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

Bedate, A.M., L.C. Herrero, and J.A. Sanz. 2009. "Economic valuation of a contemporary art museum: correction of hypothetical bias using a certainty question." *Journal of Cultural Economics* 33:185-199.

Bishop, R.C., T.A. Heberlein, and M.J. Kealy. 1983. "Contingent Valuation of Environmental Assets: Comparison with a Stimulated Market." *Nat. Resources J.* 23:619-633.

Blamey, R.K., J.W. Bennett, and M.D. Morrison. 1999. "Yea-saying in contingent valuation surveys." *Land Economics*:126-141.

Blomquist, G.C., K. Blumenschein, and M. Johannesson. 2009. "Eliciting willingness to pay without bias using follow-up certainty statements: comparisons between probably/definitely and a 10-point certainty scale." *Environmental and Resource Economics* 43:473-502.

Blumenschein, K., M. Johannesson, G.C. Blomquist, B. Liljas, and R.M. O'Conor. 1998. "Experimental results on expressed certainty and hypothetical bias in contingent valuation." *Southern Economic Journal*:169-177.

Bosworth, R., T.A. Cameron, and J. DeShazo. 2015. "Willingness to pay for public health policies to treat illnesses." *Journal of health economics* 39:74-88.

Boyce, R.R., T.C. Brown, G.H. McClelland, G. Peterson, and W.D. Schulze. 1989. "Experimental evidence of existence value in payment and compensation contexts." *Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning, Interim Report* 2.

Broadbent, C.D., J.B. Grandy, and R.P. Berrens. 2010. "Testing for hypothetical bias in a choice experiment using a local public good: riparian forest restoration." *International Journal of Ecological Economics and Statistics* 19:1-19.

Brower, L.P., O.R. Taylor, E.H. Williams, D.A. Slayback, R.R. Zubieta, and M.I. Ramirez. 2012. "Decline of monarch butterflies overwintering in Mexico: is the migratory phenomenon at risk?" *Insect Conservation and Diversity* 5:95-100.

Brown, G.M., and J.F. Shogren. 1998. "Economics of the endangered species act." *The Journal of Economic Perspectives* 12:3-20.

Brown, K.M., and L.O. Taylor. 2000. "Do as you say, say as you do: evidence on gender differences in actual and stated contributions to public goods." *Journal of Economic Behavior & Organization* 43:127-139.

Brown, T.C., P.A. Champ, R.C. Bishop, and D.W. McCollum. 1996. "Which response format reveals the truth about donations to a public good?" *Land Economics*:152-166.

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* 49:330-342.

Burton, A.C., K.S. Carson, S.M. Chilton, and W.G. Hutchinson. 2007. "Resolving questions about bias in real and hypothetical referenda." *Environmental and Resource Economics* 38:513-525.

Byrnes, B., C. Jones, and S. Goodman. 1999. "Contingent valuation and real economic commitments: evidence from electric utility green pricing programmes." *Journal of Environmental Planning and Management* 42:149-166.

Cameron, T.A., and J. DeShazo. 2013. "Demand for health risk reductions." *Journal of Environmental Economics and Management* 65:87-109.

Cameron, T.A., G.L. Poe, R.G. Ethier, and W.D. Schulze. 2002. "Alternative non-market value-elicitation methods: are the underlying preferences the same?" *Journal of Environmental Economics and Management* 44:391-425.

Campbell, D., R. Scarpa, and W.G. Hutchinson. 2008. "Assessing the spatial dependence of welfare estimates obtained from discrete choice experiments." *Letters in Spatial and Resource Sciences* 1:117-126.

Carlsson, F., P. Frykblom, and C.J. Lagerkvist. 2005. "Using cheap talk as a test of validity in choice experiments." *Economics Letters* 89:147-152.
Carlsson, F., M. Kataria, A. Krupnick, E. Lampi, Å. Löfgren, 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 & Organization* 89:105-121.
Carson, R.T., and M. Czajkowski (2013) "A new baseline model for estimating willingness to pay from discrete choice models." In *International choice modelling conference, Sydney.*
Carson, R.T., N.E. Flores, K.M. Martin, and J.L. Wright. 1996. "Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods." *Land Economics* 72:80-99.
Carson, R.T., and T. Groves. 2007. "Incentive and informational properties of preference questions." *Environmental and Resource Economics* 37:181-210.
Carson, R.T., T. Groves, and J.A. List. 2014. "Consequentiality: A theoretical and experimental exploration of a single binary choice." *Journal of the Association of Environmental and Resource Economists* 1:171-207.
Champ, P.A., and R.C. Bishop. 2006. "Is willingness to pay for a public good sensitive to the elicitation format?" *Land Economics* 82:162-173.
Champ, P.A., R.C. Bishop, T.C. Brown, and D.W. McCollum. 1997. "Using donation mechanisms to value nonuse benefits from public goods." *Journal of Environmental Economics and Management* 33:151-162.
Champ, P.A., R. Moore, and R.C. Bishop. 2009. "A comparison of approaches to mitigate hypothetical bias." *Agricultural & Resource Economics Review* 38:166-180.
Christie, M., and C.D. Azevedo. 2009. "Testing the consistency between standard contingent valuation, repeated contingent valuation and choice experiments." *Journal of Agricultural Economics* 60:154-170.
Clark, J. 2002. "House money effects in public good experiments." *Experimental Economics* 5:223-231.
Clucas, B., K. McHugh, and T. Caro. 2008. "Flagship species on covers of US conservation and nature magazines." *Biodiversity and Conservation* 17:1517-1528.
Collins, J.P., and C.A. Vossler. 2009. "Incentive compatibility tests of choice experiment value elicitation questions." *Journal of Environmental Economics and Management* 58:226-235.
Corrigan, J.R., D.P.T. Depositario, R.M. Nayga, X. Wu, and T.P. Laude. 2009. "Comparing open-ended choice experiments and experimental auctions: An application to golden rice." *American Journal of Agricultural Economics* 91:837-853.
Cummings, R.G., and L.O. Taylor. 1998. "Does realism matter in contingent valuation surveys?" *Land Economics*:203-215.
Cummings, R.G., and L.O. Taylor. 1999. "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method." *American Economic Review* 89:649-665.

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* 95:1136-1154.

del Saz Salazar, S., and L.G. Menéndez. 2007. "Estimating the non-market benefits of an urban park: Does proximity matter?" *Land Use Policy* 24:296-305.

Diffendorfer, J.E., J.B. Loomis, L. Ries, K. Oberhauser, L. Lopez-Hoffman, D. Semmens, B. Semmens, B. Butterfield, K. Bagstad, and J. Goldstein. 2014. "National Valuation of Monarch Butterflies Indicates an Untapped Potential for Incentive-Based Conservation." *Conservation Letters* 7:253-262.

Donfouet, H.P.P., E. Makaudze, P.-A. Mahieu, and E. Malin. 2011. "The determinants of the willingness-to-pay for community-based prepayment scheme in rural Cameroon." *International journal of health care finance and economics* 11:209-220.

Drichoutis, A.C., A. Vassilopoulos, J.L. Lusk, and R.M. Nayga Jr (2015) "Reference dependence, consequentiality and social desirability in value elicitation: A study of fair labor labeling." In *143rd Joint EAAE/AAEA Seminar, March 25-27, 2015, Naples, Italy.* European Association of Agricultural Economists.

Ducarme, F., G.M. Luque, and F. Courchamp. 2013. "What are “charismatic species” for conservation biologists." *BioSciences Master Reviews* 10:1-8.

Ehmke, M.D., J.L. Lusk, and J.A. List. 2008. "Is hypothetical bias a universal phenomenon? A multinational investigation." *Land Economics* 84:489-500.

Ethier, R.G., G.L. Poe, W.D. Schulze, and J. Clark. 2000. "A comparison of hypothetical phone and mail contingent valuation responses for green-pricing electricity programs." *Land Economics* 76:54-67.

Fiebig, D.G., M.P. Keane, J. Louviere, and N. Wasi. 2010. "The generalized multinomial logit model: accounting for scale and coefficient heterogeneity." *Marketing Science* 29:393-421.

Fifer, S.J. 2011. "Hypothetical bias in stated preference experiments: is it a problem? And if so, how do we deal with it?" PhD Dissertation, University of Sydney.

Fix, P., and J. Loomis. 1998. "Comparing the economic value of mountain biking estimated using revealed and stated preference." *Journal of Environmental Planning and Management* 41:227-236.

Flockhart, D., J.B. Pichancourt, D.R. Norris, and T.G. Martin. 2015. "Unravelling the annual cycle in a migratory animal: breeding-season habitat loss drives population declines of monarch butterflies." *Journal of Animal Ecology* 84:155-165.

Foster, V., I.J. Bateman, and D. Harley. 1997. "Real and Hypothetical Willingness to Pay For Environmental Preservation: A Non-Experimental Comparison." *Journal of Agricultural Economics* 48:123-138.

Fox, J.A., J.F. Shogren, D.J. Hayes, and J.B. Kliebenstein. 1998. "CVM-X: calibrating contingent values with experimental auction markets." *American Journal of Agricultural Economics* 80:455-465.

Gibbard, A. 1973. "Manipulation of voting schemes: a general result." *Econometrica: journal of the Econometric Society*:587-601.
Gracia, A., M.L. Loureiro, and R.M. Nayga. 2011. "Are valuations from nonhypothetical choice experiments different from those of experimental auctions?" *American Journal of Agricultural Economics* 93:1358-1373.
Greene, W.H., and D.A. Hensher. 2010. "Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models." *Transportation* 37:413-428.
Groothuis, P.A., T.M. Mohr, J.C. Whitehead, and K. Cockerill. 2017. "Endogenous Consequentiality in Stated Preference Referendum Data: The Influence of the Randomly Assigned Tax Amount." *Land Economics* 93:258-268.
Haab, T.C., and K.E. McConnell. 2002. *Valuing environmental and natural resources: the econometrics of non-market valuation*: Edward Elgar Publishing.
Hanley, N., F. Schläpfer, and J. Spurgeon. 2003. "Aggregating the benefits of environmental improvements: distance-decay functions for use and non-use values." *Journal of Environmental Management* 68:297-304.
Harrison, G.W., and E.E. Rutström (2008) "Experimental evidence on the existence of hypothetical bias in value elicitation methods." In C. Plott, and V. Smith eds. *Handbook of experimental economics results.* Elsevier, pp. 752-767.
Hensher, D., N. Shore, and K. Train. 2005. "Households' willingness to pay for water service attributes." *Environmental and Resource Economics* 32:509-531.
Herriges, J., C.L. Kling, C. Liu, and J. Tobias. 2010. "What are the consequences of consequentiality?" *Journal of Environmental Economics and Management* 59:67-81.
Hidano, N., T. Kato, and K. Izumi (2005) "Reciprocity, consequentiality and willingness-to-pay in contingent valuation: An experimental panel analysis on climate changes." In *European Association of Environmental and Resource Economics.* Bremen, Germany.
Hoehn, J.P., and A. Randall. 1987. "A satisfactory benefit cost indicator from contingent valuation." *Journal of Environmental Economics and Management* 14:226-247.
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* 42:445-469.
Horowitz, J.K., and K.E. McConnell. 2002. "A review of WTA/WTP studies." *Journal of Environmental Economics and Management* 44:426-447.
Hwang, J., D.R. Petrolia, and M.G. Interis. 2014. "Consequentiality and Opt-out Responses in Stated Preference Surveys." *Agricultural and Resource Economics Review* 43:471.

Interis, M., and D. Petrolia. 2014. "The effects of consequentiality in binary-and multinomial-choice surveys." *Journal of Agricultural and Resource Economics* 39:201-216.
Jacquemet, N., R.-V. Joule, S. Luchini, and J.F. Shogren. 2013. "Preference elicitation under oath." *Journal of Environmental Economics and Management* 65:110-132.
Jepsen, S., D.F. Schweitzer, B. Young, N. Sears, M. Ormes, and S.H. Black. "Conservation status and ecology of monarchs in the united states." NatureServe and Xerces Society for Invertebrate Conservation.
Jianjun, J. "An Experiment on Hypothetical and Real Willingness to Pay." Economy and Environment Program for Southeast Asia.
Johannesson, M., B. Liljas, and P.-O. Johansson. 1998. "An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions." *Applied Economics* 30:643-647.
Johnston, R.J., and J.M. Duke. 2009. "Willingness to pay for land preservation across states and jurisdictional scale: implications for benefit transfer." *Land Economics* 85:217-237.
Johnston, R.J., D. Jarvis, K. Wallmo, and D.K. Lew. 2015. "Multiscale Spatial Pattern in Nonuse Willingness to Pay: Applications to Threatened and Endangered Marine Species." *Land Economics* 91:739-761.
Johnston, R.J., and M. Ramachandran. 2014. "Modeling spatial patchiness and hot spots in stated preference willingness to pay." *Environmental and Resource Economics* 59:363-387.
Juutinen, A., Y. Mitani, E. Mäntymaa, Y. Shoji, P. Siikamäki, and R. Svento. 2011. "Combining ecological and recreational aspects in national park management: A choice experiment application." *Ecological economics* 70:1231-1239.
Kang, M.J., A. Rangel, M. Camus, and C.F. Camerer. 2011. "Hypothetical and real choice differentially activate common valuation areas." *The Journal of Neuroscience* 31:461-468.
Kaufman, V. (2014) "Service Initiates Status Review of Monarch Butterfly under the Endangered Species Act." In *Newsroom.* US Fish & Wildlife Service.
Kemper, N., R.M. Nayga Jr, J. Popp, and C. Bazzani (2016) "The Effects of Honesty Oath and Consequentiality in Choice Experiments." In *2016 Annual Meeting, July 31-August 2, 2016, Boston, Massachusetts.* Agricultural and Applied Economics Association.
Kimenju, S.C., H. De Groote, and U.B. Morawetz (2006) "Comparing accuracy and costs of revealed and stated preferences: the case of consumer acceptance of yellow maize in East Africa." In *Contributed paper for presentation at the International Association of Agricultural Economists conference, Australia.*
Krinsky, I., and A.L. Robb. 1986. "On approximating the statistical properties of elasticities." *The Review of Economics and Statistics*:715-719.
Ladenburg, J., J.O. Dahlgaard, and O. Bonnichsen. 2011. "Testing the effect of a short cheap talk script in choice experiments." *Nationaloekonomisk Tidsskrift* 149:25-54.

Ladenburg, J., and S.B. Olsen. 2014. "Augmenting short cheap talk scripts with a repeated opt-out reminder in choice experiment surveys." *Resource and Energy Economics* 37:39-63.

Landry, C.E., and J.A. List. 2007. "Using ex ante approaches to obtain credible signals for value in contingent markets: evidence from the field." *American Journal of Agricultural Economics* 89:420-429.

León, C.J., J.E. Araña, J. de León, and M.M. González. 2016. "The economic benefits of reducing the environmental effects of landfills: heterogeneous distance decay effects." *Environmental and Resource Economics* 63:193-218.

Lewis, K.E., C. Grebitus, and R.M. Nayga. 2016. "US consumers' preferences for imported and genetically modified sugar: Examining policy consequentiality in a choice experiment." *Journal of Behavioral and Experimental Economics* 65:1-8.

Li, X., K.L. Jensen, C.D. Clark, and D.M. Lambert. 2016. "Consumer willingness to pay for beef grown using climate friendly production practices." *Food Policy* 64:93-106.

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

List, J.A. 2003. "Using random nth price auctions to value non-market goods and services." *Journal of Regulatory Economics* 23:193-205.

List, J.A., and C.A. Gallet. 2001. "What experimental protocol influence disparities between actual and hypothetical stated values?" *Environmental and Resource Economics* 20:241-254.

List, J.A., M. Margolis, and J.F. Shogren. 1998. "Hypothetical–actual bid calibration of a multigood auction." *Economics letters* 60:263-268.

List, J.A., and J.F. Shogren. 1998. "Calibration of the difference between actual and hypothetical valuations in a field experiment." *Journal of Economic Behavior & Organization* 37:193-205.

List, J.A., and J.F. Shogren. 2002. "Calibration of willingness-to-accept." *Journal of Environmental Economics and Management* 43:219-233.

Little, J., and R. Berrens. 2004. "Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis." *Economics Bulletin* 3:1-13.

Little, J., C.D. Broadbent, and R.P. Berrens. 2012. "Meta-Analysis of the Probability of Disparity Between Actual and Hypothetical Valuation Responses: Extension and Preliminary New Results." *Western Economics Forum* 11:1-12.

Loomis, J. 2011. "What's to know about hypothetical bias in stated preference valuation studies?" *Journal of Economic Surveys* 25:363-370.

Loomis, J., T. Brown, B. Lucero, and G. Peterson. 1996. "Improving validity experiments of contingent valuation methods: results of efforts to reduce the disparity of hypothetical and actual willingness to pay." *Land Economics*:450-461.

Loomis, J., A. Gonzalez-Caban, and R. Gregory. 1994. "Do reminders of substitutes and budget constraints influence contingent valuation estimates?" *Land Economics* 70:499-506.
Loomis, J., and L. Santiago. 2013. "Economic Valuation of Beach Quality Improvements: Comparing Incremental Attribute Values Estimated from Two Stated Preference Valuation Methods." *Coastal Management* 41:75-86.
Loomis, J., K. Traynor, and T. Brown. 1999. "Trichotomous choice: a possible solution to dual response objectives in dichotomous choice contingent valuation questions." *Journal of Agricultural and Resource Economics* 24:572-583.
Loomis, J.B. 2014. "Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys." *Journal of Agricultural and Resource Economics* 39:34-46.
Loomis, J.B., and D.S. White. 1996. "Economic benefits of rare and endangered species: summary and meta-analysis." *Ecological Economics* 18:197-206.
Loureiro, M.L., W.J. Umberger, and S. Hine. 2003. "Testing the initial endowment effect in experimental auctions." *Applied Economics Letters* 10:271-275.
Lusk, J.L., and T.C. Schroeder. 2006. "Auction bids and shopping choices." *Advances in Economic Analysis & Policy* 6.
Macmillan, D.C., T.S. Smart, and A.P. Thorburn. 1999. "A field experiment involving cash and hypothetical charitable donations." *Environmental and Resource Economics* 14:399-412.
Mahieu, P.-A. 2010. "Does gender matter when using cheap talk in contingent valuation studies." *Economics Bulletin* 30:2955-2961.
Mahieu, P.-A., P. Riera, and M. Giergiczny. 2012. "The influence of cheap talk on willingness-to-pay ranges: some empirical evidence from a contingent valuation study." *Journal of environmental planning and management* 55:753-763.
Maynard, L.J., J.G. Hartell, A.L. Meyer, and J. Hao. 2004. "An experimental approach to valuing new differentiated products." *Agricultural Economics* 31:317-325.
McFadden, D. 1973. "Conditional logit analysis of qualitative choice behavior."
Meas, T., W. Hu, M. Batte, T. Woods, and S. Ernst. 2015. "Substitutes or Complements? Consumer Preferences for Local and Organic Food Attributes." *American Journal of Agricultural Economics* 97:1044-1071.
Metrick, A., and M.L. Weitzman. 1996. "Patterns of behavior in endangered species preservation." *Land Economics*:1-16.
Meyerhoff, J. 2013. "Do turbines in the vicinity of respondents' residences influence choices among programmes for future wind power generation?" *Journal of choice modelling* 7:58-71.
Meyerhoff, J., and U. Liebe. 2009. "Status quo effect in choice experiments: empirical evidence on attitudes and choice task complexity." *Land Economics* 85:515-528.

Michaud, C., D. Llerena, and I. Joly. 2013. "Willingness to pay for environmental attributes of non-food agricultural products: a real choice experiment." *European Review of Agricultural Economics* 40:313-329.
Mitani, Y., and N. Flores. 2007. "Does gender matter for demand revelation in threshold public goods experiments." *Econ Bull* 3:1-7.
Mitani, Y., and N.E. Flores. 2014. "Hypothetical bias reconsidered: payment and provision uncertainties in a threshold provision mechanism." *Environmental and Resource Economics* 59:433-454.
Mitchell, R.C., and R.T. Carson. 1989. *Using surveys to value public goods: the contingent valuation method*: Resources for the Future.
Mjelde, J.W., Y.H. Jin, C.-K. Lee, T.-K. Kim, and S.-Y. Han. 2012. "Development of a bias ratio to examine factors influencing hypothetical bias." *Journal of environmental management* 95:39-48.
Morrison, M., and T.C. Brown. 2009. "Testing the effectiveness of certainty scales, cheap talk, and dissonance-minimization in reducing hypothetical bias in contingent valuation studies." *Environmental and Resource Economics* 44:307-326.
Mozumder, P., and R.P. Berrens. 2007. "Investigating hypothetical bias: induced-value tests of the referendum voting mechanism with uncertainty." *Applied Economics Letters* 14:705-709.
Murphy, J.J., P.G. Allen, T.H. Stevens, and D. Weatherhead. 2005. "A meta-analysis of hypothetical bias in stated preference valuation." *Environmental and Resource Economics* 30:313-325.
Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison, and T. McGuckin. 1994. "Hypothetical surveys and real economic commitments." *Land Economics* 70:145-154.
Nelson, J.P., and P.E. Kennedy. 2009. "The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment." *Environmental and Resource Economics* 42:345-377.
Nepal, M., R.P. Berrens, and A.K. Bohara. 2009. "Assessing perceived consequentiality: Evidence from a contingent valuation survey on global climate change." *International Journal of Ecological Economics and Statistics™* 14:14-29.
Onwujekwe, O., K. Hanson, and J. Fox-Rushby. 2005. "Do divergences between stated and actual willingness to pay signify the existence of bias in contingent valuation surveys?" *Social science & medicine* 60:525-536.
Pandit, R., V. Subroy, S.T. Garnett, K.K. Zander, and D. Pannell (2015) "A review of non-market valuation studies of threatened species and ecological communities." In *Report to the National Environmental Science Programme, Department of the Environment, Canberra.*
Pate, J., and J. Loomis. 1997. "The effect of distance on willingness to pay values: a case study of wetlands and salmon in California." *Ecological Economics* 20:199-207.
Pleasants, J.M., and K.S. Oberhauser. 2013. "Milkweed loss in agricultural fields because of herbicide use: effect on the monarch butterfly population." *Insect Conservation and Diversity* 6:135-144.

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* 23:105-131.
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* 87:353-365.
Poe, G.L., and C.A. Vossler. 2011. *Consequentiality and contingent values: an emerging paradigm*: Edward Elgar Publishing: Cheltenham, UK.
Pollinator Health Task Force. 2015. *National Strategy to Promote the Health of Honey Bees and Other Pollinators.*
Polomé, P. 2003. "Experimental evidence on deliberate misrepresentation in referendum contingent valuation." *Journal of Economic Behavior & Organization* 52:387-401.
Ready, R.C., J.C. Buzby, and D. Hu. 1996. "Differences between continuous and discrete contingent value estimates." *Land Economics* 72:397-411.
Ready, R.C., P.A. Champ, and J.L. Lawton. 2010. "Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment." *Land Economics* 86:363-381.
Ritland, D.B., and L.P. Brower. 1991. "The viceroy butterfly is not a Batesian mimic." *Nature* 350:497-498.
Rolfe, J., and J. Windle. 2012. "Distance decay functions for iconic assets: assessing national values to protect the health of the Great Barrier Reef in Australia." *Environmental and Resource Economics* 53:347-365.
Satterthwaite, M.A. 1975. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of economic theory* 10:187-217.
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* 90:994-1010.
Schaafsma, M., R. Brouwer, and J. Rose. 2012. "Directional heterogeneity in WTP models for environmental valuation." *Ecological Economics* 79:21-31.
Schulze, W., G. McClelland, D. Waldman, and J. Lazo. 1996. "Sources of bias in contingent valuation." *The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs, edited by D. Bjornstad and J. Kahn. Cheltenham, UK: Edward Elgar*.
Schulze, W.D., R.C. d'Arge, and D.S. Brookshire. 1981. "Valuing environmental commodities: some recent experiments." *Land Economics* 57:151-172.
Shrestha, R.K., and J.B. Loomis. 2001. "Testing a meta-analysis model for benefit transfer in international outdoor recreation." *Ecological economics* 39:67-83.
Silva, A., R.M. Nayga Jr, B.L. Campbell, and J.L. Park. 2011. "Revisiting cheap talk with new evidence from a field experiment." *Journal of Agricultural and Resource Economics* 36:280-291.
Stachtiaris, S., A. Drichoutis, R. Nayga, and S. Klonaris (2011) "Can religious priming induce truthful preference revelation?" In *Working Paper.*

Stanley, T., H. Doucouliagos, M. Giles, J.H. Heckemeyer, R.J. Johnston, P. Laroche, J.P. Nelson, M. Paldam, J. Poot, and G. Pugh. 2013. "Meta-analysis of economics research reporting guidelines." *Journal of Economic Surveys* 27:390-394.

Stanley, T.D. 2008. "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and statistics* 70:103-127.

Sutherland, R.J., and R.G. Walsh. 1985. "Effect of distance on the preservation value of water quality." *Land Economics* 61:281-291.

Taylor, L.O., M. McKee, S.K. Laury, and R.G. Cummings. 2001. "Induced-value tests of the referendum voting mechanism." *Economics letters* 71:61-65.

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

Tunçel, T., and J.K. Hammitt. 2014. "A new meta-analysis on the WTP/WTA disparity." *Journal of Environmental Economics and Management* 68:175-187.

US Fish and Wildlife Service (2016) "Species Status Assessment: Monarch Butterfly." In., US Fish and Wildlife Service.

Van Bueren, M., and J. Bennett. 2004. "Towards the development of a transferable set of value estimates for environmental attributes." *Australian Journal of Agricultural and Resource Economics* 48:1-32.

Vásquez, W.F., and D. Franceschi. 2013. "System reliability and water service decentralization: Investigating household preferences in Nicaragua." *Water resources management* 27:4913-4926.

Vásquez, W.F., P. Mozumder, J. Hernández-Arce, and R.P. Berrens. 2009. "Willingness to pay for safe drinking water: Evidence from Parral, Mexico." *Journal of Environmental Management* 90:3391-3400.

Veisten, K., and S. Navrud. 2006. "Contingent valuation and actual payment for voluntarily provided passive-use values: Assessing the effect of an induced truth-telling mechanism and elicitation formats." *Applied Economics* 38:735-756.

Veríssimo, D., I. Fraser, J. Groombridge, R. Bristol, and D.C. MacMillan. 2009. "Birds as tourism flagship species: a case study of tropical islands." *Animal Conservation* 12:549-558.

Vlaev, I. 2012. "How different are real and hypothetical decisions? Overestimation, contrast and assimilation in social interaction." *Journal of Economic Psychology* 33:963-972.

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

Vossler, C.A., M. Doyon, and D. Rondeau. 2012. "Truth in consequentiality: theory and field evidence on discrete choice experiments." *American Economic Journal: Microeconomics*:145-171.

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

Vossler, C.A., and M. McKee. 2006. "Induced-value tests of contingent valuation elicitation mechanisms." *Environmental and Resource Economics* 35:137-168.

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

Walpole, M.J., and N. Leader-Williams. 2002. "Tourism and flagship species in conservation." *Biodiversity & Conservation* 11:543-547.

Weaver, R., and D. Prelec. 2013. "Creating truth-telling incentives with the Bayesian truth serum." *Journal of Marketing Research* 50:289-302.

Welsh, M.P., and G.L. Poe. 1998. "Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach." *Journal of Environmental Economics and Management* 36:170-185.

Whitehead, J.C., and T.L. 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.

Williams, M. 2013. "Consumer Willingness to Pay for Transitional Organic Produce."

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.

VITA

Jerrod Penn

Education

- MS, Agricultural Economics, 2013
  Thesis: Valuation of Recreational Beach Quality and Water Quality Management Strategies In Oahu
  University of Kentucky, Lexington, KY
  Major Professor: Wuyang Hu
- MS, Economics, University of Kentucky, Lexington, KY December 2016
- BA, Economics *Suma Cum Laude*, 2010
- BA, Political Science *Cum Laude*, 2010
  - Minor in Food and Resource Economics and Business Administration
  - Political Science Certificate in International Relations

  University of Florida, Gainesville, FL
  Undergraduate Advisor: Dr. Lisa House lahouse@ufl.edu

Research Specialties

- Primary: Environmental and Resource Economics
  - Policy and Management, Nonmarket Valuation, Recreation, Species Conservation
- Secondary: Marketing and Consumer Economics
  - Sustainability and Food Marketing, Hospitality, Economic Issues within Entomology

Employment and Academic Service

- Lecturer. Agricultural, Environmental and Development Economics, The Ohio State University, Columbus, OH, Fall 2014.
- Research Intern, Economic Research Service, US Department of Agriculture, Washington D.C., Summer 2015.
- Graduate Research Assistant. Agricultural Economics, Univ. of Kentucky, Lexington, KY, August 2010-Present.
- Graduate Teaching Assistant. Agricultural Economics, Univ. of Kentucky, Lexington, KY, August 2012-Present.
- Academic Quiz Bowl Advisor, Agricultural Economics, Univ. of Kentucky, Lexington, KY, January 2011-Present
- Teaching Assistant, Food and Resource Economics, Univ. of Florida, Gainesville, FL, Summer 2010
- Research Intern, Food and Resource Economics Dept., Univ. of Florida, Gainesville, FL, Summer 2009 and Summer 2010

Academic Awards

- Endowed Doctoral Fellowship Award from the University of Kentucky Association of Emeriti Faculty, 2017, $2,500 (University).
- 1st Place Quiz Bowl Team as Advisor/Coach, 2016, AAEA, Boston, MA (National).
- Teaching, Learning and Communication Section's Graduate Teaching Award, 2015 AAEA Meeting in San Francisco, CA (National).
- 2nd Place Quiz Bowl Team as Advisor/Coach, 2015, AAEA, San Francisco, CA (National).
- Univ. of Kentucky Provost's Outstanding Teaching Assistant Award, 2015 (University).
- 2nd Place (UC-Davis 1st and MIT 3rd) in the United States Assoc. for Energy Economics (USAEE) 2014 Case Study Competition with M Lang, W Martin, D Nichols, and E Rutledge. New York City, NY (National).
- Travel Scholarships, AAEA and Univ. of Kentucky Graduate School, 2012-2016. $1,870 cumulative.
- Daniel R. Reedy Quality Achievement Award, Univ. of Kentucky Graduate School 2012-13, 2013-14, & 2014-2015. $9,000 cumulative (University).
- Academic Excellence Scholarship, Univ. of Kentucky Student Government, 2013, $1,000 (University).
- Gamma Sigma Delta Outstanding Master's Student Award, Univ. of Kentucky College of Agriculture 2012 (College).
- 1st Place Quiz Bowl Team, 2010 AAEA Meeting, Denver, CO, as contestant (National).
- 2nd Place Quiz Bowl Team, 2010 SAEA Meeting, Orlando, FL, as contestant (Regional).
- 2nd Place Best Chapter, Univ. of Florida National Agri-Marketing Assoc. (NAMA) Team, 2010 NAMA Meeting in Kansas City, MO (National).

Funding

- Hu, W, Meyer AL, and J Penn. $13,500, 2016. UK Student Survey of Local Beef Consumption and Sustainability Initiatives. Jointly supported by the UK Food Connection and Student Sustainability Council.
- Support Assistance to faculty for Resiliency in Transitioning Economies, $241,000, October 2016 submitted to USDA-NIFA National Needs Fellows (Declined). Grant to fund four graduate students trained in "Targeted Expertise Shortage Areas."
- Penn J, Zumdick S, Crowley P, and W Hu. $18,884, 2016. Supporting Monarchs and Pollinators through Citizen Science and Public Engagement. Univ. of Kentucky 2015 Sustainability Challenge Grants/Student Sustainability Council.
- Penn J, Hu W, and M Potter. $15,500, 2014. Protect-A-Bed Corporation. Economic Study of Bed Bugs in the Hospitality Industry.
- Penn J, and L Maynard. $500, 2014. Dept. of Agricultural Economics' Centennial Quiz Bowl Invitational. Univ. of Kentucky College of Agriculture, Food and Environment's annual Barnhart Fund for Excellence Award.

- Penn J, and W Hu. $7,855, 2014. Univ. of Kentucky Arboretum Project. Lexington-Fayette Arboretum and Student Sustainability Council.
- Straathof D, J Penn, and W Hu. $2,000, 2013. Multi-University Comparison of Electricity Output in Fitness Centers from ReRev Technology. Univ. of Kentucky's Student Sustainability Council.
- Penn J, and W Hu. $6,320, 2013. Faculty and Staff Sustainability Survey. Univ. of Kentucky's Student Sustainability Council.
- Penn J, and Y Kusonose. $750, 2012. Research Activity Award for International Web Seminar Series. Univ. of Kentucky's College of Agriculture.
- Penn J, A Hancock, R Lee, and W Hu. $2,890 (Grant and In-kind), 2012. Johnson Fitness Center, Dining Services, and Student Sustainability Council, all at the Univ. of Kentucky.
- Penn J, W Hu, and A McLaughlin. $5,460, 2012. Student Sustainability Survey. Univ. of Kentucky's Student Sustainability Council.
- Penn J, W Hu, and A McLaughlin. $1,000, 2011. Service Learning Mini-Grant. Univ. of Kentucky's Office of Undergraduate Education.
- Penn J, and J Schieffer. $500, 2011. Research Activity Award for Web Seminar Series. Univ. of Kentucky's College of Agriculture.
- Penn J. $2,000 Scholarship, 2010. Mechanized Citrus Harvesting Feasibility. Food and Resource Economics Dept., Univ. of Florida.

Peer-Reviewed Publications

- **Penn J,** and W Hu. Understanding Hypothetical Bias: An Enhanced Meta-Analysis. *Conditionally Accepted, American Journal of Agricultural Economics*.
- Zhong H, Hu W and **J Penn**. Farmers' Willingness and Expected Economic Benefit to Adopt BMPs: An Application of Multivariate Imputation by Chained Equation Method. *Conditionally Accepted, Journal of Agricultural and Resource Economics.*
- **Penn J**, and HM Sandberg. Agricultural and Resource Economics Ph.D. Students: Who are They and What Do They Want? *Accepted, in final preparation, NACTA (North American Colleges and Teachers of Agriculture) Journal.*
- **Penn J**, Penn H, Potter M, and W Hu. 2017. Bed Bugs and Hotels: Traveler Insights and Implications for the Industry. *American Entomologist,* 63(2) 79-88*.*
- **Penn J**, W Hu, L Cox, and L Kozloff. 2016. Economic Implications of Non-Point Source Water Pollution in Hawaii Tourism. *Marine Resource Economics*. 31(1) 47-62.
- Hu W, Qing P, **Penn J**, Pelton M, and A Pagaloulatos. 2015. Rider Preferences and Economic Values for Equestrian Trails. *Journal of Environmental Planning and Management*. 58(7): 1154-1172.

- **Penn J**, Hu W, Cox L, and L Kozloff. 2014. Resident and Tourist Preferences for Stormwater Management Strategies in Oahu, Hawaii. *Ocean and Coastal Management.* 98(1): 79-85.
- Hancock A*, **Penn J**, and W Hu. 2014. Augmenting Electricity Output of Ellipticals through Behavioral Change. *Sustainability: The Journal of Record*. 7(5): 255-261.
- **Penn J**, Brown D, and L Maynard. 2014. The Economic Impact of Bed Bugs within the Hospitality Industry. *Consortium Journal of Hospitality & Tourism*. 19(1): 22-47.

Proceedings and Other Publications

- Potter M, **Penn J**, and W Hu. Bed Bugs, Hotels, and Travelers: Attitudes and Implications. Proc. Intl' Conf. Urban Pests. Birmingham, UK. *In Press*.
- **Penn J**, Hu W, and M Potter. 2015. Disturbed! *Pest Control Technology*. November 2015 issue.

Working Papers

- **Penn J**, and W Hu. The Effect of Forced Choice with Constant Choice Experiment Complexity. *Revise and Resubmit*.
- **Penn J**, and W Hu. Euthanizing Value of a Statistical Life: Monetizing Differences in Public Perception and Alternatives. *Revise and Resubmit*.
- **Penn J**, Penn H, and W Hu. Public knowledge of monarch conservation in Kentucky. *Under Review*.
- Trull N*, **Penn J**, and W Hu. Public Support for Growth and Funding in Built Environments: Case of an Arboretum. *Under Review*.
- **Penn J,** and W Hu. Nudging against Hypothetical Bias by Defaulting to the Opt-Out in an Online Choice Experiment. *Under Review*.
- **Penn J**, and W Hu. The Presence of Hypothetical Bias within Spatial Decay and Charismatic Species: An Application of Monarch and Viceroy Butterflies. *Working Paper Available.*
- **Penn J**, and W Hu. A Comparison of Ex Ante and Ex Post Consequentiality Efficacy to Reduce Hypothetical Bias. *Working Paper Available.*
- **Penn J**, and W Hu. Making the Most of Cheap Talk in an Online Survey. *Working Paper Available.*
- **Penn J**, and W Hu. Determinants of Cheap Talk Efficacy: A Meta-Analysis
- **Penn J**, and W Hu. Hypothetical Bias from Willingness to Accept Questions: Is it Bias in the Outliers?

Presentations

- Gabrielson Z, Penn J, and S Zumdick. Student-Led Pollinator Habitat Restoration Project: Low-Cost Implementation and Potential Applications. Accepted poster, Assoc. for the Advancement of Sustainability in Higher Education (AASHE) 2017, San Antonio, TX.
- Penn J, and W Hu. 2017. Accounting for Hypothetical Bias for Space and Species- A Case Study of Monarch and Viceroy Butterflies. Accepted presentation, AAEA, Chicago, IL.
- Penn J, and W Hu. 2017. Hypothetical Bias from Willingness to Accept Questions: Is it Bias in the Outliers? Invited track presentation, AAEA, Chicago, IL.
- Zuo N and J Penn 2017. Teaching as a Graduate Student: A One-credit Teaching Module Case. Accepted poster, NACTA 2017, Purdue University, West Lafayette, IN.
- Penn H, J Penn, and W Hu. 2016. Public Attitudes on Monarch Conservation. Ohio Valley Entomological Association (OVEA), West Lafayette, IN.
- Hart R, Penn J, and S Zumdick. Experiences and Strategies of a Student-Run Massive Monarch Restoration on Off-campus Sites. Accepted Student Summit Case Study, AASHE 2016, Baltimore, MD.
- Penn J, and W Hu. 2016. Making the Most of Cheap Talk in an Online Survey. Accepted presentation, AAEA, Boston, MA.
- Marshall K, MK Adjemian, J Penn, and T Hubbs. 2016. Just How Local is the Local Corn Basis? A Historical Decomposition of US Corn Prices. Accepted presentation, AAEA, Boston, MA.
- Penn J, and W Hu. 2016. Euthanizing Value of a Statistical Life: Monetizing differences in public perception and alternatives. Accepted presentation, W3133. Portland, OR.
- Teaching Tips from AAEA Award Winners. 2015. Organized session, AAEA, San Francisco, CA.
- Penn J, and W Hu. 2015. Revisiting Determinants of Hypothetical Bias: An Up-To-Date Meta-Analysis. Accepted presentation, W3133, Pensacola, FL.
- Vassalos M, J Penn, and K Davidson. 2015. Organizer/Participant Graduate Student Training: How to be an Effective Teaching Assistant. Accepted organized symposium, SAEA, Atlanta, GA.
- Penn J, W Hu, and N Trull. 2015. Public support for growth and funding in Built Environments: Case of an Arboretum. Accepted presentation, SAEA, Atlanta, GA.
- Penn J, W Hu, L Cox, and L Kozloff. 2014. The Effect of Forced Choice with Constant Choice Experiment Complexity. Accepted presentation, AAEA, Minneapolis, MN.
- Penn J, W Hu, L Cox, and L Kozloff. 2014. Resident and Tourist Preferences for Stormwater Management Strategies in Oahu. Accepted presentation, W3133, Orange Beach, AL.

- Penn, J, HM Sandberg, G Ferro, B McFadden, and N Nayaupane. 2014. Agricultural Economics Graduate Training: Distinguishing the Expectation, Effort, and Experience to Succeed in Master's or Ph.D. Programs. Accepted organized symposium, SAEA, Dallas, TX.
- Penn, J, W Hu, and L Maynard. 2013. "Undergraduate Learning Through Research and On-Campus Consulting: A Multi-Course Experience." Invited track presentation, AAEA, Washington DC.
- Penn J, and HM Sandberg. 2013. "The Attitudes and Expectations of Graduate Students in Agricultural Economics: A National Survey." Accepted presentation, AAEA, Washington DC.
- Hu W, J Penn, and L Cox. 2013. "Enhancing Conjoint Analysis with Respondents' Self-Constructed Preferred Alternative: Application in Choices of Dolphin Excursions." Accepted presentation, SAEA, Orlando, FL.
- Penn J, HM Sandberg, B Barnett, D Leatham, and M Wetzstein. 2013. Organizer, Moderator, and Participant of "A Matter of Opinion: The perspective and experience of Graduate Students versus Graduate Coordinators in Agricultural Economics." Accepted organized symposium, SAEA, Orlando, FL.
- Penn J, L Maynard, and D Brown. 2012. "Bed Bug Anxiety: Travelers' Willingness to Pay to Avoid Them." Presentation, Entomological Society of America meeting, Knoxville, TN.
- Penn J, W Hu, L Cox, and L Kozloff. 2012. Resident and Tourist Preferences for Stormwater Management Strategies in Hawaii with a Cost-Benefit Analysis." Accepted poster, AAEA 2012, Seattle, WA.
- Penn J, J Gillespie, HM Sandberg, and L Kompaniyets. 2012. Organizer, Moderator, and Participant of "Graduate Degrees in Agricultural Economics: Expectation and Preparation." Accepted organized symposium, SAEA, Birmingham, AL.
- Penn J, W Hu, L Cox, and L Kozloff. 2012. "Beach Quality and Recreational Values: A Pictorial Stated Preference Analysis." Accepted presentation, SAEA, Birmingham, AL.
- Penn J, D Staley, C Smith, and S Saghaian. 2011. "Advertising Content Analysis of Online Children's Television Programming" Accepted poster, AAEA, Pittsburgh, PA.
- Penn, J, A Matopoulos, and L House. 2010. "Response to Out of Stock Produce and its Underlying Economic Considerations." Accepted presentation, SAEA, Orlando, FL.

Invited Presentations

- "Human Economic Behavior and Values of Environmental Amenities" with Wuyang Hu. 2016. Beijing Normal University, Beijing, China, Dec 12.
- "The Extent of Valuation Hypothetical Bias in Species and Space: Comparing Monarch and Viceroy Butterflies" with Wuyang Hu. 2016. China Agricultural University, Beijing, China, Dec 13. Renmin University of China, Beijing, China, Dec 14.

- "Travelers' thoughts on Bed Bugs" with Michael F. Potter and Ken Haynes. 2015. 45th Annual Univ. of Kentucky Pest Control Short Course. Lexington, KY, Nov 3-5.
- "Hotels, Travelers, and Bed Bugs—Oh my!" with Wuyang Hu and Michael F. Potter. 2015. National Pest Management Association Global Bed Bug Summit, Denver, CO. Jan 7-9.

Advising (presentations/achievements available upon request)

- Adam Hancock
- Nathanial U. Trull
- Jason Simon
- Daniella Straathof
- Zoe Gabrielson
- Josey Moore

Reviewer

- Ecological Economics (1)
- Canadian Journal of Agricultural Economics (2)
- European Review of Agricultural Economics (1)
- Journal of Agricultural and Resource Economics (1)
- Journal of Agricultural and Applied Economics (1)
- Advances in Health Economics and Health Services Research

Professional Memberships

- Agricultural and Applied Economics Association
- Southern Agricultural Economics Association
- Northeastern Agricultural and Resource Economics Association
- Association of Environmental and Resource Economists
- Entomology Society of America
- National Pest Management Association

Other Academic Services

Student Sustainability Council, Univ. of Kentucky
Director of Development (2012-2013, 2015-16), Member (2011-2017), Intern Advisor (2015-2017), Univ. of Kentucky

President's Sustainability Advisory Committee, Univ. of Kentucky
Student Representative (2011-2014, 2016-17)

Agricultural Economics Graduate Student Organization, Univ. of Kentucky
President 2013-14, Vice-President 2012-13, Treasurer 2011-12 & 2015-16

Software Experience

- Statistical: Stata, NLogit/LimDep, R, SAS, JMP
- Mapping: ArcGIS (portfolio of work available upon request)

List of References

Wuyang Hu, H.B. Price Professor
Wuyang.hu@uky.edu, 859-257-7271
Agricultural Economics, University of Kentucky

Brian Roe, McCormick Professor
roe.30@osu.edu, 614-688-5777
Agricultural, Environmental & Development, Economics, The Ohio State University

Leigh Maynard, Chair
leigh.maynard@uky.edu, 859-257-7277
Agricultural Economics, University of Kentucky

Michael Adjemian, Economist
madjemian@ers.usda.gov, 202-694-5576
Economic Research Service, USDA