Advances in Economic Analysis & Policy

Volume 6, *Issue* 2 2006 *Article* 2

FIELD EXPERIMENTS

Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments

John A. List* Paramita Sinha†

Michael H. Taylor‡

*University of Chicago, Jlist@uchicago.edu

†University of Maryland, College Park, psinha@arec.umd.edu

‡University of Maryland, College Park, mtaylor@arec.umd.edu

Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments*

John A. List, Paramita Sinha, and Michael H. Taylor

Abstract

Critics of stated preference methods argue that hypothetical bias precludes survey techniques from providing reliable economic values for non-market goods and services, rendering estimation of the total economic benefits of public programs fruitless. This paper explores a relatively new methodology to obtain the total value of non-market goods and services—choice experiments—which conveniently provide information on the purchase decision as well as the characteristic value vector. The empirical work revolves around examining behavior in two very different field settings. In the first field study, we explore hypothetical bias in the purchase decision by eliciting contributions for a threshold public good in an actual capital campaign. To extend the analysis a level deeper, in a second field experiment we examine both the purchase decision and the marginal value vector via inspection of consumption decisions in an actual marketplace. In support of the new valuation design, both field experiments provide some evidence that hypothetical choice experiments combined with "cheap talk" can yield credible estimates of the purchase decision. Furthermore, we find no evidence of hypothetical bias when estimating marginal attribute values. Yet, we do find that the "cheap talk" component might induce internal inconsistency of subjects' preferences in the choice experiment.

KEYWORDS: field experiments, valuing nonmarketed goods

---

*Correspondence: John A. List, Professor, The University of Chicago, Department of Economics 1126 East 59th Street, Chicago, IL 60636; email: jlist@uchicago.edu; website: http://www.arec.umd.edu/jlist/ We would like to thank Don Fullerton and three anonymous reviewers for very sound suggestions that lead to a fundamental revision of this study. Richard Carson, Rachel Croson, Glenn Harrison, Daniel Millimet, and Laura Taylor also provided useful suggestions. Seminar participants at the University of Chicago, Harvard University, University of Maryland, and the University of Wyoming provided helpful insights. The usual caveats apply.

Benefit-cost analysis remains the central paradigm used throughout the public sector. A fundamental issue in properly estimating the total benefits of non-market goods and services is whether hypothetical statements are a credible indicator of actual preferences. Although practitioners often assume that behavior in contingent scenarios is closely associated with actual market behavior, many skeptics remain within the academic and policy communities (see Diamond and Hausman's (1994) critical review of the contingent valuation method (CVM)). The potential bias is particularly troublesome given that an accurate value estimate for non-market goods and services is necessary to advance efficient public policies. While CVM remains the most popular method to estimate the total economic value of the commodity in question, unfortunately it has not been the acme of perfection: in practice, many studies find that hypothetical statements in CVM markets exceed actual values.[1]

In this study, we explore a methodology that is relatively new to economics to estimate values for non-market goods and services: choice experiments. A choice experiment (CE) asks subjects to choose between scenarios that are described by attributes of the good and therefore the standard discrete choice CVM is a specific type of choice experiment. A CE conveniently combines Lancaster's (1966) characteristics theory of value with random utility theory (McFadden, 1974). Although CE has firm roots theoretically, to gain widespread acceptability it must be scrutinized under the same rigorous tests to which other CVM institutions have been subjected (e.g., Cummings et al., 1997; Cummings and Taylor, 1999; List, 2001). While these other institutions have generally not performed well, in that hypothetical and actual behavior has not perfectly matched, this study represents a first attempt in the field to provide a firm understanding of the external validity properties of the CE approach.

Under the CE approach, it is important to recognize that hypothetical bias can occur at two levels: i) the decision to purchase and ii) the intra-buy decision (i.e., conditional on purchasing, the marginal value vector might be biased). Thus, a CE has several distinct advantages over more traditional CV surveys. For example, they provide a natural manner in which to estimate the value of attributes of an environmental good. This may be important if decision makers are concerned with changing attribute levels. CVM, on the other hand, can only address questions concerning gaining or losing the good as a whole. Under some

[1] As List et al. (2004) note, CVM is a set of survey-based approaches for eliciting Hicksian compensating or equivalent surplus values for a hypothetical change in a good or program. While the CVM approach is practically quite important, critics contend that hypothetical bias severely limits credibility. In the validation study literature, scholars have attempted to discern the degree of hypothetical bias by comparing hypothetical and actual statements of value in experimental markets, where the actual value is assumed to represent *true* preferences. The interested reader can read the debate between Cummings et al. (1997), Haab et al. (1999), and Smith (1999).

CE designs, the problem posed by List et al. (2004) that is commonly encountered in dichotomous choice CVM models is absent, since respondents choose between alternatives and are not faced with an "all or nothing" choice. Since many tests of scope are naturally part of CE, embedding problems that arise in CVM might be avoided. CE can also be used to obtain marginal values of attributes that may be difficult to identify using revealed preference data due to co-linearity problems or lack of sufficient variation. This may assist in benefit transfer as well, if socioeconomic variables are included in the model.

To begin the empirical exploration, we examine whether the hypothetical CE method can elicit true purchase decisions for non-marketed goods and services. To do so, we carry-out a new natural field experiment that includes mailing 3,000 solicitations to individual households asking for contributions to the Center for Environmental Policy Analysis (CEPA) at the University of Central Florida (UCF). The 3,000 Central Florida residents were randomly assigned to three different groups of 1,000, with each group asked to fund a computer for use at CEPA. Following the received literature, the "real" treatment subjects were actually asked to contribute funds to CEPA, while in the hypothetical and hypothetical with "cheap talk" treatments, subjects were asked a hypothetical contribution question.[2] Insights gained from these treatments reveal the potential for a hypothetical CE to provide a distribution of responses that is similar to a CE with monetary incentives: responses across the real and hypothetical with "cheap talk" treatments are not statistically different from each other.[3]

Since one nice characteristic of the CE approach is its natural ability to estimate the marginal value vector, to explore a level deeper it is important to consider whether CE can elicit meaningful marginal attribute values. A formal test of this second level of hypothetical bias requires an examination of a market where the good's value is critically linked to its attributes. In this sense, the sportscard market represents a reasonable choice since a sportscard's value is quite sensitive to its characteristics. We make use of comparative static changes along two dimensions of more than 200 Nolan Ryan sportscards (valued at $15-$50)—front picture centering and corner fraying of the card—to explore behavior in a framed field experiment (see Harrison and List, 2004). We again find that responses across the real and hypothetical with "cheap talk" treatments are not statistically different from each other. In terms of the intra-buy decision, we find little evidence of estimated marginal value differences across the hypothetical and

[2] The use of the term "cheap talk" in this study differs from the use of the term in the game theory literature. In this study "cheap talk" refers to an ex-ante method of attenuating hypothetical bias where the subject of hypothetical bias is made an integral part of the CVM questionnaire. This usage of the term "cheap talk" is consistent with Cummings and Taylor (1999).

[3] The astute reader will note that the "real" treatment may not convey subjects' "real" valuations because of the free-rider problem. This issue is absent in our private good treatments.

real regimes. Furthermore, this finding is robust across both inexperienced consumers and those who have intense experience; though there is some evidence that hypothetical bias remains among the most experienced consumers.

While this might be viewed as good news for the CE method, the data in the sportscard field experiment are sufficiently rich to perform a third validity test: violations of preference consistency. Given that our choice experiment permits an analysis of the internal consistency of a subject's preferences, we explore this aspect of decision making across the three treatments. We find that the tendency to make internally consistent decisions is more pronounced among experienced subjects. An unexpected result, however, is that subjects in the hypothetical with cheap talk treatment are more likely to make inconsistent decisions compared to subjects in the other two treatments. This is because they are not adjusting their behavior on the margin; rather, subjects in the cheap talk treatment are refusing purchase in an arbitrary manner. This finding has important implications if it is robust to other goods and other settings.

Taken as a whole, these results should be considered mixed news for the CE approach. On the one hand, the CE method combined with "cheap talk" can elicit credible market signals. Additionally, there appears to be little evidence of hypothetical bias in the marginal valuation estimates. While other CVM institutions have not universally induced consistent behavior across contingent and actual regimes, the CE method may achieve external validity through its ability to structure the value formation process in a meaningful and familiar way. Indeed, this intuition is consistent with evidence in Huber et al. (2002), who find that subjects view choice decisions as *i*) the most realistic tasks, and *ii*) the mode in which they feel most confident making decisions. On the other hand, if cheap talk induces subjects to violate other fundamental economic assumptions, through introducing new biases into subject's decision making, care should be taken when applying this method and much more research is necessary before one can properly advocate its use to policymakers.

The remainder of this study proceeds as follows. Section II provides a short literature survey placing our study in proper perspective, reviews the general theory underlying the CE mechanism, and summarizes a few recent studies that implement the CE approach. Sections III and IV present the experimental designs and empirical results of the two field experiments. Section V concludes.

1. Previous Literature and Choice Experiments in Practice

Understanding whether people over- or under-state their actual preferences for a non-marketed good when asked a hypothetical question remains an important issue in current policy debates. This importance follows from President Clinton's Executive Order 12866, which reaffirms an earlier executive order from the

Reagan Administration requiring that federal agencies consider costs, benefits, and economic impacts of regulations prior to their implementation. In terms of computing benefit estimates, policymakers understand that the flexible and holistic contingent valuation approach makes it the "only game in town" in a wide variety of situations.

Early work in the hypothetical bias literature suggests that people tend to overstate their real willingness to pay in hypothetical markets. In response, the National Oceanic and Atmospheric Administration's (NOAA) blue-ribbon panel recommends that hypothetical bids be deflated using a "divide by 2" rule unless these bids can be calibrated using real market data (NOAA, 1994, 1996). The NOAA rule has triggered a search for a calibration function to correct systematic bias between intentions and actions in valuation exercises.

Recently, several studies have reviewed the vast literature on the observed relationship between real and hypothetical values.[4] Rather than repeat those reviews here, we provide a brief summary. As List and Gallet (2001) note, the research began with Bohm's (1972) seminal experimental lab study which compared bids in hypothetical and real experimental markets that elicited subjects' stated value to sneak preview a Swedish television show. His results suggest people moderately overstate their real values when asked a hypothetical question. Subsequent lab research has generally supported Bohm's initial findings.

List and Gallet (2001) take a step back from the burgeoning literature and use a meta-analysis to determine whether important experimental parameters systematically affect the relationship between hypothetical and real responses. Using 174 sets of results from 29 papers, they find that the various CVM elicitation methods, such as open-ended valuation, dichotomous choice questions, Vickrey second price auctions, Smith auctions, the BDM (Becker-DeGroot-Marschak) approach, and random *n*th price auctions can importantly influence the observed differences between hypothetical and real statements of value. This finding opens up the possibility that there might be certain elicitation institutions that yield less hypothetical bias than others. While these other valuation institutions have been subjected to careful tests (e.g., Cummings et al., 1997), much less time has been spent evaluating choice experiments.

1.1 Choice Experiments in Practice

Many practitioners recently have used binary discrete choice elicitation schemes to estimate total economic values, perhaps due to the endorsement of the NOAA

[4] See Foster et al. (1997) for a non-experimental comparison of real and hypothetical willingness to pay (WTP) statements and List and Gallet (2001) and List and Shogren (2002) who update the Foster et al. (1997) study.

Panel (Arrow et al., 1993). The basis of the Panel's recommendation rests on the notion that binary questions have the well-known property of being incentive-compatible in many circumstances.

Choice experiments are an extension of the binary discrete choice institution, but have several advantages. For instance, CE usually makes use of many choice pairs and allows individuals multiple opportunities to express their preferences in a manner that might allow attenuation of the issues stressed in List et al. (2004). Further, CE can be represented by a simple utilitarian paradigm. One such model is a discrete choice variant based on random utility theory. Under this approach, the CE model makes the assumption that individual i will select option j if expected utility (u_{ij}) exceeds the expected utility (u_{ik}) for all alternative $k=1,...,K$ choices, implying that the probability of selecting an alternative increases as the utility associated with its selection increases. Accordingly, utility that an individual derives from an alternative can be considered to be a function of the attributes of that alternative. As such, a simple model of utility can be represented by a deterministic and a random component:

$$u_{ij} = \beta' X_j + e_{ij}, \tag{1}$$

where X_j is a vector of observable attributes, β is a vector of estimated parameters, and e_{ij} is the random error component. A well-known property of equation (1) is that if the e_{ij} follow a Weibull distribution and are independently and identically distributed, then the probability that individual i will select alternative j is

$$P_{ij} = exp(z\beta' X_j) / \sum_{k=1}^{K} \exp(z\beta' X_k), \tag{2}$$

where z is a scale parameter usually assumed to equal 1.

1.2 Previous CE Studies and Cheap Talk

While the choice-based method is relatively new in the area of environmental economics, some published research has used the structural framework outlined above as a platform to examine choices in surveys.[5] A few recent examples are found in Adamowicz et al. (1994; 1998). In the first study, the authors use data from hypothetical choice statements on attributes of water-based recreation sites to investigate underlying preferences. Using mail-survey data, Adamowicz et al. (1994) collect information from over 413 individuals who provided choices across 13 attributes that could reasonably affect utility during recreational visits. Empirical results from a random utility model suggest that CE data, when combined with revealed preference data, have the capability of providing

[5] See Louviere et al. (2000) for a discussion of broader uses of CE.

information not easily captured in other survey designs. The authors conclude by noting that CE "may offer a valuable alternative to contingent valuation" (p. 290).

In the 1998 study, Adamowicz and associates present the first CE investigation into passive use values by examining hypothetical choices concerning the protection of old growth forests in west central Alberta. The CE questionnaire presented alternative woodland designs described in five attributes, which each took one of four levels. Using data from a mail survey completed by 447 Edmonton, Canada, residents, the authors found credible value estimates across each attribute dimension, as measures of convergent validity were obtained that suggested the CE survey performed well.

As aforementioned, in our study the term "cheap talk" refers to an ex-ante method of eliminating hypothetical bias. The effectiveness of cheap talk at attenuating hypothetical bias has been shown to be robust, but context dependent. For example, it has been shown that cheap talk is successful at eliminating hypothetical bias in a referendum for public goods (Cummings and Taylor, 1999) but ineffective with more experienced market participants (List, 2001; Aadland and Caplan, 2003). Further, cheap talk may only be effective for higher payments (Murphy et al., 2005) and shorter cheap talk scripts yield mixed evidence (see, e.g., Poe et al. 2002; Aadland and Caplan, 2003). In this study we employ cheap talk scripts of two lengths: a shorter script was used in our field experiment where we solicited donations for a Center of Environmental Policy Analysis at the University of Central Florida (p.9), while a longer script was used in our second field experiment involving sportscard purchasing decisions (Appendix C).

2. External Field Test of Choice Experiments: Public Goods

As previously noted, although CE are increasing in popularity and have considerably expanded the boundaries of non-market valuation, they have yet to be tested with the same rigor as other valuation institutions.[6] In the spirit of previous work, the following experiments represent a first test of the CE approach in the field. As noted earlier, in the burgeoning validation study literature, scholars have attempted to discern the degree of hypothetical bias by comparing hypothetical and real statements of value in experimental markets. In general,

[6] A first consideration is whether the mechanism can work in practice. Hence, to begin the analysis, we examine both hypothetical and real choices within an induced-value laboratory experiment. Gathering data from 150 University of Arizona undergraduate students, we find that while some individual mistakes are observed, in aggregate the mechanism is demand-revealing in both hypothetical and real treatments. In terms of assessing the mechanism credibility, these results represent a step forward since they suggest that if individuals have well-behaved preferences, the CE method can elicit truthful responses. These results are available upon request. A laboratory experiment that examines whether hypothetical choices are similar to actual choices is due to Carlsson and Martinsson (2001).

these studies reveal that typically respondents overstate their true preferences in hypothetical exercises.

Unlike these other approaches, it is important to recognize that under the CE method two distinct types of hypothetical bias can emerge. First, it could be the case that hypothetical bias is present in the purchase decision. Second, even if hypothetical bias is not present in the purchase decision (i.e., agents in the hypothetical scenario are equally likely to purchase the good as agents in the actual scenario), it might be present in the estimates of the marginal values of the attributes.

To provide initial insights into the first type of bias in a public goods setting, we took advantage of a unique opportunity we were provided at UCF. While a faculty member at UCF, one of the coauthors was approached to spearhead a capital campaign to fund a new Center for Environmental Policy Analysis (CEPA). Having received permission to design the fundraising campaign as an experiment, we split the full capital campaign into several smaller capital campaigns, each of which would serve as a separate experimental treatment. We then solicited contributions from 3,000 Central Florida residents, asking them to fund a computer for use at CEPA. The 3000 subjects were randomly assigned into one of three treatment groups of 1000 subjects: real, hypothetical, and hypothetical with cheap talk. While List and Lucking-Reiley (2002) also use CEPA for their fundraiser, this is the first use of these data in the literature.

In carrying out the field experiment, we solicited donors in a way that closely matched current fundraising standards. With advice from fundraising companies *Donnelley Marketing* (and associates) in Englewood, Colorado, and *Caldwell* in Atlanta, Georgia, we followed generally accepted rules believed to maximize overall contributions. First, we purchased names and addresses of households in the Central Florida area that met two criteria: *i*) annual household income greater than $70,000, and *ii*) the household had given to a charity in the previous year. From *Donnelley Marketing*, we purchased the names and home addresses of 3,000 Central Floridians who met both criteria.

Second, we designed an attractive brochure describing the new center and its purpose. Excerpts from the brochure read as follows:

> The primary objective of The *Center for Environmental Policy Analysis* (CEPA) will be to improve the quality of Florida's public and private decisions that have environmental, economic, and resource-use implications. In addition, the CEPA will propose economically efficient solutions to national and international problems ranging from endangered species protection to global issues such as climate change and sustainable development.

> The CEPA will accomplish these tasks through an integrated program of communications, publications, and education, designed to lead from awareness through knowledge to action. Through these programs, the CEPA will improve communication between the public, including various governmental branches, and the business community.
>
> The CEPA will also offer courses, seminars, and an opportunity for students to conduct research under the guidance of some of the nation's leading scholars in environmental and resource economics. The CEPA's current faculty have served on government advisory bodies, editorial boards, and have been visiting scholars at prominent universities around the globe.

Third, we constructed a personalized letter of solicitation that noted CEPA's role within the Central Florida community, the total funds required to purchase the computer, and the number of solicitations sent out (1,000 in each treatment). We also explained that contributions in excess of the amount required for the computer would be used for other purposes at CEPA, noted the tax deductibility of the contribution, and closed the letter with contact information in case the potential donors had questions.

The text of the solicitation letter was identical across the three treatments, except for the use of subjunctive language in the hypothetical and hypothetical with cheap talk treatments, and the inclusion of the cheap talk script in the hypothetical with cheap talk treatment. Also, at this point it is useful to be clear about two aspects of the solicitation letters. First, we explicitly noted in the letter that the mechanism was "closed" in that subjects should not believe that they could "free-ride" off of contributions to CEPA made by others outside the group: "If we fail to raise the $3,000 from this group of 1,000 individuals, we will not be able to purchase the computer….".[7] Second, to make the solicitation as natural as possible, we asked only one choice-based question in the letter (this excerpt is taken from the letter sent to subjects in the real treatment): "We would like you to consider making a contribution towards the purchase of a $3,000 computer to be used by researchers at CEPA by answering the choice question below:"

[7] This inclusion, coupled with the results from the induced value experiment described earlier, should help to ensure that the actual contributions represented a useful benchmark with which to compare the hypothetical choices.

Please circle your choice of Column 1, 2, or 3:

Column 1	*Column 2*	*Column 3*
$20 donation	*$5 donation*	*Status quo: no donation*
UCF contributes $25	*UCF contributes $5*	*UCF contributes $0*

For comparability, we used the same choice question in each of the 3,000 solicitations. A composite version of the solicitation letter from the actual treatment can be found in Appendix A. All letters were mailed between January 29 and February 3, 2000. The shortened cheap talk script is contained in Appendix A.

Empirical results for the public good field treatments are presented in Table 1. Of first note are the response rates in each treatment. Of the 1,000 mailed surveys in each treatment, 105 (10.5 percent), 118 (11.8 percent), and 95 (9.5 percent) surveys were returned in the real, hypothetical, and hypothetical with cheap talk treatments. These figures are in the range of other CE field surveys—e.g., Adamowicz et al. (1994) began with a sample of 4,497 and received 413 (9.2 percent) accurately completed surveys.[8] The real treatment raised $310 for CEPA, whereas $780 was pledged in the hypothetical, and $365 in the hypothetical with cheap talk treatment.[9]

In certain respects, the data in Table 1 suggest that hypothetical decisions do not consistently map well into real behavior. For example, in the hypothetical treatment, 29 and 40 subjects hypothetically gave $20 and $5, respectively. This does not compare well to the 8 and 30 donors that chose (and therefore sent in their personal checks or cash) to give these amounts in the real treatment. Indeed, using a Pearson chi-square test of homogeneity of distributions, which is

[8] In fairness, however, they mailed out only 785 surveys (they mailed surveys only to subjects who agreed on the telephone to complete the survey). List and Lucking-Reiley (2002), using data not reported herein, test whether seed money can induce higher contribution rates, and they report an even lower response rate in their fundraiser. As they note, however, this level of giving was surprisingly high. Mixer (1993) indicates in a fundraising handbook that direct-mail solicitations typically yield a response rate of only about 1% from "cold" lists of potential donors. This discrepancy is most likely due to the fact that our mailing list included some "warm" list donors who had given to UCF previously.

[9] Unfortunately, after subtracting the costs of the mailing list, materials, postage, and labor, the net result was a loss of more than $3,000. Before proceeding to discuss the empirical results, it is worthwhile to mention that upon receiving the returned surveys, we cross-checked for differences across demographic characteristics of participants in each of the three survey types. An ANOVA test indicates that the respective samples for the three treatments do not differ by the following socioeconomic characteristics: income, age, gender of head-of-household, and number of kids. Nevertheless, sample selection might still be a cause for concern so the reader should take care when making inference from the results below. For example, consistent with any experimental study that involves selection into the experiment one should be cautious and understand that the findings are a representative description for the population conforming to the selection criteria.

distributed as χ^2 with 2 degrees of freedom, we find that the likelihood of choosing to donate is related to which of these two regimes, real and hypothetical, the subject was randomly allocated—$\chi^2 = 15.44$. This corresponds to rejecting the null of homogeneity at p-value < 0.01.

Table 1: Public Good Experimental Results

	Column Choice			
Regime	$20	$5	$0	Total
Real	8	30	67	105
Hypothetical	29	40	49	118
Hypothetical with cheap talk	12	25	58	95

Note: Figures represent the number of subjects that chose that particular column. For example, "8" in row 1 column 1 means that 8 people in the actual treatment chose to contribute $20.

A similar comparison between subjects in the real and hypothetical with cheap talk treatments yields different insights: the cheap talk script induced donors to submit responses that were quite similar to choices in the real treatment. This is most evident via examination of the percentages of donors that chose each of the contribution columns: in the hypothetical with cheap talk treatment, 12 and 25 subjects hypothetically give $20 and $5, which is similar to column choices in the real treatment (8 and 30). Using a Pearson chi-square test of homogeneity of the real and hypothetical with cheap talk distributions, we find that the homogeneity null cannot be rejected at conventional levels - χ^2 (2 df) = 1.41 (p-value > 0.10). This result suggests that the cheap talk script combined with the CE method has the potential to induce responses that are consonant with real responses in the public good domain. Given the close relationship that this particular question has with CVM dichotomous choice questions, this result should not be surprising in light of the findings in the literature.

As elaborated on more fully in previous studies (e.g., Cummings and Taylor, 1999), one potential explanation for the success of the cheap talk script is that as subjects become aware of the potential influence of the context of a hypothetical decision on their valuation of a good, they attempt to "correct" for the hypothetical nature of the exercise. Through the internal correction process, the individual commits cognitive effort to retrieve a more accurate value for the

good in question. This general direction of findings within economics is nicely validated within social psychology, where researchers argue that subjects' judgments in various scenarios are "primed" differently—where the context of choice plays an important role in the "priming" process. In Section IV, we present evidence that cheap talk's effectiveness *might* be attributable to an introduction of new biases into a subject's decision making that counteract hypothetical bias rather than by inducing "corrective" behavior by making subject's aware of the hypothetical context of their decisions.

3. External Field Test of Choice Experiments: Private Goods

One important advantage of the CE method is its ability to estimate values of a good's attributes in a theoretically straightforward manner. Analyses of the above spirit are important in exploring one type of hypothetical bias, but an important test of the mechanism is to examine its external validity in a market where subjects are comfortable making valuation decisions over attributes of the good in question. A naturally occurring market where the good's attributes play an integral role in its valuation is the sportscard market. Sportscards represent a good choice for such an external test because any sportscard consumer readily recognizes that values are critically linked to card characteristics. This link is perhaps best illustrated by an example. Making use of the grading system of a well-known third party, Professional Sports Authenticators (PSA),[10] one could consider the value difference between a Ken Griffey Jr. 1989 *Upper Deck* baseball card graded "Gem Mint 10" by PSA, and an identical Griffey card (identical but for a slight fraying on one corner) graded "Mint 9" by PSA. The Gem Mint version of the Griffey Jr. card readily sells for $2700, whereas the identical Griffey Jr. card with slight corner fraying will garner only between $125 and $350.

This natural variation in value due to the card's quality allows us to make use of comparative static changes along several dimensions. In this exercise, we make use of Nolan Ryan 1982 *Topps* sportscards that have substantial variation across two important characteristics—front picture centering and corner fraying—to test for the presence of hypothetical bias in the marginal attribute vector. To gather an appropriate card stock, a PSA representative helped us hand-grade more than 200 Nolan Ryan 1982 *Topps* sportscards to provide enough cards that varied

[10] PSA, which grades cards on a 1-10 scale (10 being the best card), is the world's largest sportscard grading and authentication service, currently accepting over 300,000 trading cards per month for grading. PSA is the most widely accepted grading standard in the sportscard market, and maintains co-branded grading and authentication sites with *eBay*, *Yahoo!* and other online commerce and content sites. The first card ever graded by PSA was the famous T206 Honus Wagner card that recently sold for $1.1 million.

only marginally across the two attributes of interest.[11] We should note that a Gem Mint PSA graded 10 card is one that has no corner fraying and nearly perfect front centering—60/40 or better (where 50/50 is terminology for a perfectly centered sportscard). Hence, a card's grade begins at Gem Mint 10 and decreases with, for example, frayed corners or an off-centered picture (centering worse than 60/40—e.g., 70/30, 75/25, 80/20, 85/15, 90/10).[12] The cards we used in the experimental treatments were PSA graded 7-9, and each retails for $15-$50.

The field treatments, which were conducted at a sportscard show in Tucson, Arizona, are most similar (but not identical to) in methodology to List (2001). Each treatment (hypothetical, hypothetical with cheap talk, or real) had 4 steps:

Step 1. As a potential subject enters the sportscard show, the monitor inquires about his/her interest in participating in an experiment. If the individual accepts the invitation, then the monitor provides a thorough description of the sportscard on the table, ensuring that there is no ambiguity about the card's qualities.

Step 2. After being informed of the experimental instructions (see Appendix B) and whether the choice is hypothetical or real, each subject chooses Column 1, Column 2, or Column 3 for each of the six questions. The six questions are presented in Appendix B, and the following represents question 1:

1. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
80/20+FC	70/30+FC	Status quo: no
at most 1 corner frayed	at most 2 corners frayed	purchase
$5 payment	$5 payment	

where 80/20+FC represents a card that has no worse than 80/20 front picture centering. Note that we were careful to choose goods that could not be purchased in the market for $5 or less.

We should stress that these are preference-based questions and therefore in general there is no "incorrect" answer. Rather we are interested in whether hypothetical and real choices differ. But question 5 might well be considered a rationality check: both elements of the Nolan Ryan card described in column 1 are in better condition than of the card described in Column 2—better centering

[11] To acquire the Nolan Ryan cards, we opened a case of 1982 *Topps* trading cards valued at $1500 and purchased the remaining necessary cards on the open market.

[12] Centering is determined by comparing the measurements of the borders from left to right and top to bottom. The card's "centering grade" is the percentage difference at the most off-center part of the card.

and less corner wear. Hence, conditional on demanding the Nolan Ryan card, each subject should choose Column 1 in question 5.

Step 3. A six-sided die is rolled and the number appearing determines which of the six questions is actually executed.[13]

Step 4. Each subject then i) departs the experiment with the card after he/she pays $5, or ii) departs without receiving (or paying for) the card if he/she opted for the status quo or was allocated into one of the hypothetical treatments.

At this point, it is worthwhile to mention that, consistent with the CEPA field experiment, besides using the typical hypothetical treatment where the only change to the experimental instructions in Appendix B is the use of subjunctive language (e.g., *if* this were actually occurring; you *would* pay $5), we used the longer "cheap talk" scheme proposed in Cummings and Taylor (1999) to induce truthful responses. The additional "cheap talk" language is very similar to List (2001), and can be found in Appendix C.

Given that the cheap talk design has not universally worked across all subject pools (List, 2001), we also experimented with bidder type, conducting some of the experiments with professional card dealers and others with nondealers. In each case, we randomize participants into the three question formats (real, hypothetical, hypothetical with cheap talk) to ensure that treatment populations are similar. For example, while gathering the nondealer data, at the top of each hour we changed the treatment type. For dealers, we gathered data before the show began and alternated treatment type. No subjects participated in more than one treatment. And, in the real treatments, subjects used their own resources since we did not provide a participation fee.

3.1 Experimental Results

The sportscard field experiment naturally allows agents to make two decisions for each question: whether to purchase the card and, conditional on purchasing, which card to choose. Unconditional empirical results summarizing the first of these decisions are contained in Tables 2A and 2B. Table 2A presents the nondealer data, and includes responses from 225 subjects, equally distributed across the three treatments. Table 2B contains the dealer data, which include roughly 50 dealers in each treatment. Recall that only one of the questions has a

[13] Note that this procedure makes any one question binding with 1/6 probability in the actual treatments. We also ran an actual treatment in which subjects answered only one question. The data are not significantly different from the actual choice data presented below. This finding is potentially very important since it suggests that if subjects are uncertain about the monetary consequences of their decisions, they behave as if their actions will be binding. This result highlights the notion that practitioners should attempt to make the valuation question consequential (see, e.g., Carson et al., 2000).

choice that may be considered optimal if a subject demands the Nolan Ryan card at a price of $5—in question 5, the subject should *definitely* choose column 1. The remaining five questions rely on the preference trade-off between corner wear and picture centering. In Table 2A, we find that regardless of treatment type, nondealers seem to choose the "correct" choice in question 5.

In questions 2 and 5, the choices are very consistent across treatment, and using a Pearson chi-square test of homogeneity of distributions, which is distributed as χ^2 with 2 degrees of freedom, we find that the likelihood of choosing column 1, 2, or 3 is not related to which regime the subject was randomly allocated (question 2: real versus hypothetical: $\chi^2 = 1.07$; real versus hypothetical with cheap talk: $\chi^2 = 1.29$; question 5: real versus hypothetical: $\chi^2 = 2.87$; real vs. hypothetical with cheap talk: $\chi^2 = 2.69$). None of these chi-square values permit rejection of the homogeneous null at conventional significance levels.

While these results suggest that hypothetical statements might be good indicators of real preferences, findings across the other four questions are not as promising. For example, in all four questions we reject the homogeneous null between the real and hypothetical choices at the $p < .10$ level using a Pearson chi-square test (#1: $\chi^2 = 5.05$; #3: $\chi^2 = 8.18$; #4: $\chi^2 = 6.70$; #6: $\chi^2 = 4.27$). A rejection of the null in each case follows mainly because in the hypothetical treatments subjects chose to purchase the Ryan card much more often than subjects chose to purchase the card in the real treatment. This discrepancy is similar to a strand of previous results in that subjects tend to overstate their real preferences when making hypothetical statements (e.g., Cummings et al., 1995).

Nevertheless, the homogeneity null cannot be rejected at conventional significance levels when comparing responses across the real and hypothetical with cheap talk treatments. This latter result is consonant with Cummings and Taylor (1999) and List (2001), and suggests that by making hypothetical bias an integral part of the choice survey, the researcher can move the subject toward truthtelling.

Moving to the dealer data presented in Table 2B, we should first note a preliminary finding: dealers tend to have higher demands for the Nolan Ryan card at the $5 price than nondealers. This can be seen by comparing the percentage of subjects who opted for the status quo of no purchase—column 3. In the nondealer treatments, fewer than 50 percent of subjects typically purchased the card in each of the six questions (column 3), whereas in the dealer treatments at least 50 percent opted to purchase the card (columns 1 or 2) in every treatment. And, in many cases more than 75 percent of dealers chose to purchase the Nolan Ryan card.

Observed patterns in the dealer data, however, do show some similarities with data from the non-dealer sub-sample. One key similarity across each

question is that dealers in the hypothetical treatment chose to purchase the card much more often than dealers in the real treatment. Indeed, examining results from a test of homogeneity of hypothetical versus real distributions, which is again done by a Pearson chi square test (with 2 degrees of freedom), we find that the homogeneity null is rejected at the $p < .05$ level for questions #1 ($\chi^2 = 6.72$), #2 ($\chi^2 = 10.67$), #4 ($\chi^2 = 6.56$), and #6 ($\chi^2 = 6.63$), and at the $p < .10$ level for question 3 ($\chi^2 = 5.79$). The only case where hypothetical and real responses are not significantly different is question 5.

A much different outcome emerges when we compare real choices with decisions in the hypothetical with cheap talk treatment. While dealers in the hypothetical with cheap talk treatment chose to purchase the Nolan Ryan card more often than dealers in the real treatment, statistically these choices are *indistinguishable* at the $p < .05$ level for all 6 questions. Note, however, that the choices are *nearly* (statistically) different in questions 1 and 6: $\chi^2 = 5.78$ and 5.00, which are both significant at the $p < .10$ level. The direction of this "near significance" result is consistent with findings reported in List (2001), Lusk (2003), and Aadland and Caplan (2003). In particular, List (2001) notes that experienced subjects may not be easily swayed by the cheap talk design as they have a well-structured preference ordering for the good in question. The data herein suggest that the cheap talk script used in tandem with the CE approach has the potential to attenuate hypothetical bias among dealer choices.

While these results are suggestive, they can be extended by recognizing that we have panel data and therefore more powerful statistical tests can be employed. We wish to examine the value of the attributes like frayed corners, but first we need to correct for endogeneity in the choice of whether to purchase at all. In the first stage:

$$(3) \qquad Purchase_{it} = \beta' X_{it} + u_{it}$$

where $Purchase_{it}$ equals unity if agent i opted to purchase in question t, and equals zero otherwise; X_{it} includes treatment dummies—hypothetical, which equals 1 if the treatment was hypothetical and 0 otherwise, and hypothetical with

Table 2A: Sportscard Experimental Results—Nondealers

		Column Choice			
Question	Regime	1	2	3	Total
1	Real	14	11	50	75
	Hypothetical	18	20	37	75
	Hypothetical with cheap talk	11	17	47	75
2	Real	22	7	46	75
	Hypothetical	22	11	42	75
	Hypothetical with cheap talk	18	11	46	75
3	Real	11	9	55	75
	Hypothetical	20	17	38	75
	Hypothetical with cheap talk	12	15	48	75
4	Real	7	15	53	75
	Hypothetical	18	17	40	75
	Hypothetical with cheap talk	10	9	56	75
5	Real	29	2	44	75
	Hypothetical	35	5	35	75
	Hypothetical with cheap talk	38	3	34	75
6	Real	8	17	50	75
	Hypothetical	16	20	39	75
	Hypothetical with cheap talk	15	12	48	75

Note: Figures represent the number of nondealers that chose that particular column. For example, *14* in row 1 column 1 indicates that 14 nondealers chose column 1 for question 1 in the real treatment.

Table 2B: Sportscard Experimental Results—Dealers

Question	Regime	Column Choice 1	2	3	Total
1	Real	15	17	16	48
	Hypothetical	24	19	6	49
	Hypothetical with cheap talk	15	27	7	49
2	Real	28	4	16	48
	Hypothetical	31	13	5	49
	Hypothetical with cheap talk	26	10	13	49
3	Real	13	12	23	48
	Hypothetical	20	17	12	49
	Hypothetical with cheap talk	17	15	17	49
4	Real	21	13	14	48
	Hypothetical	19	24	6	49
	Hypothetical with cheap talk	23	12	13	49
5	Real	35	1	12	48
	Hypothetical	40	2	6	49
	Hypothetical with cheap talk	39	2	7	49
6	Real	17	17	15	48
	Hypothetical	19	25	5	49
	Hypothetical with cheap talk	20	23	6	49

Note: Figures represent the number of dealers that chose that particular column. For example, *15* in row 1 column 1 means that 15 people in the real treatment chose column 1 for question 1.

cheap talk, which equals 1 if the treatment was hypothetical with cheap talk and 0 otherwise. Thus, the baseline is the real treatment. Thus, for example, if the hypothetical dummy is significant, it implies that the subjects in the hypothetical treatment behave in a statistically different way from subjects in the real treatment when deciding whether to purchase. As a control for question type, we also include five dichotomous variables for the question numbers (making question 6 the baseline). In addition, we examine dealer and non-dealer data in separate regressions.

The second stage explores the second type of hypothetical bias by examining the purchaser's decisions via estimation of the following equation:

$$(4) \qquad Column1_{it} = \alpha' Z_{it} + v_{it}$$

where $Column1_{it}$ equals unity if agent i chose column 1 in question t, and equals zero otherwise; Z_{it} includes regressors of interest. In this case, we infer values by examining choices across the two attributes, centering and frayed corners (FC). To operationalize this procedure, we first define an index for both attributes such that a higher value indicates a superior attribute: centering: 80/20 centering = 1; 75/25 centering = 2; 70/30 centering = 3; 60/40 centering = 4; FC: 4 frayed corners = 1; 3 frayed corners = 2; 2 frayed corners = 3; 1 frayed corner =4.[14] We difference these index numbers across column 1 and 2 choices and include interactions of the treatment dummy variables with the differences in the attributes to test for the second type of hypothetical bias.

There could be at least one major problem with our estimating strategy as outlined above. If we estimate the two equations separately, we are making the assumption that the data include all individuals of the underlying population of interest. However, we observe column 1 only if individuals have chosen to purchase the card (i.e. individuals self-select into purchasing the card). These individuals may have certain unobserved characteristics so that the sample used to estimate equation (4) may not be representative. Thus, there is a possibility of correlated errors, u and v; in this case, estimating (3) and (4) separately will give us inconsistent estimates. We therefore run a Heckman two-step model to account for this problem.[15]

Empirical results are presented Tables 3 and 4. The top panel in each table provides summary estimates for the purchase decision. Consistent with the unconditional insights, there is hypothetical bias evident for both dealers and non-dealers in the first stage—the coefficient of the hypothetical dichotomous variable

[14] We would have preferred to use dichotomous variables for each of the attributes, but there was not enough variation in the data to enable us to achieve model convergence.

[15] In the random effects probit model, the likelihood function follows Butler and Moffitt (1982).

Table 3: Sample Selection Model for Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"

Regressors	Coefficients	Z
Hypothetical	0.65** (0.29)	2.23
Hypothetical with cheap talk	0.34 (0.24)	1.39
Question 1	-0.17 (0.09)	-1.84
Question 2	-0.21** (0.09)	-2.26
Question 3	-0.66** (0.11)	-5.89
Question 4	-0.26** (0.10)	-2.53
Question 5	-0.09 (0.11)	-0.86
Constant	0.69** (0.21)	3.24

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"

Regressors	Coefficients	Z
Difference in centering	0.33** (0.11)	2.90
Difference in FC	0.35** (0.16)	2.16
Diff in centering × hypothetical with cheap talk	-0.01 (0.09)	-0.13
Diff in centering × hypothetical	-0.09 (0.11)	-0.79
Diff in FC × hypothetical with cheap talk	-0.09 (0.13)	-0.71
Diff in FC × hypothetical	-0.03 (0.13)	-0.27
Constant	-0.13 (0.16)	-0.83

Note: This table provides results for the dealers. The independent variables for the selection equation are the treatments and the question numbers. The independent variables for the second stage are the two attributes (difference in frayed corners and differences in centering) and the attributes interacted with the treatments ("hypothetical" and "hypothetical with cheap talk"). Robust standard errors are in parentheses beneath coefficient estimates. These standard errors have been corrected for clustering. ** denotes statistical significance at the $p < .05$ level.

Table 4: Sample Selection Model for Non-Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"

Regressors	Coefficients	Z
Hypothetical	0.38** (0.19)	1.98
Hypothetical with cheap talk	0.13 (0.17)	0.76
Question 1	0.07 (0.06)	1.21
Question 2	0.03 (0.06)	0.58
Question 3	0.02 (0.09)	0.20
Question 4	-0.12 (0.08)	-1.57
Question 5	0.33** (0.07)	4.66
Constant	-0.48** (0.13)	-3.69

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"

Regressors	Coefficients	Z
Difference in centering	0.18** (0.08)	2.10
Difference in FC	0.35** (0.12)	2.89
Diff in centering × hypothetical with cheap talk	0.08 (0.08)	1.01
Diff in centering × hypothetical	-0.02 (0.07)	-0.23
Diff in FC × hypothetical with cheap talk	-0.16 (0.11)	-1.48
Diff in FC × hypothetical	-0.18 (0.11)	-1.55
Constant	0.66 (0.36)	1.84

Note: This table provides results for the non-dealers. The independent variables for the selection equation are the treatments and the question numbers. The independent variables for the second stage are the two attributes (difference in frayed corners and differences in centering) and the attributes interacted with the treatments ("hypothetical" and "hypothetical with cheap talk"). Robust standard errors are in parentheses beneath coefficient estimates. These standard errors have been corrected for clustering. ** denotes statistical significance at the $p < .05$ level.

is positive and statistically significant at the $p < .05$ level. In the case of dealers, the coefficient leads to a marginal effect estimate of roughly 0.18 , indicating that dealers are 18 percent more likely to "purchase" than subjects in the real treatment. Similarly, for non-dealers, the marginal effect estimate is about 0.15, indicating that subjects are 15 percent more likely to "purchase" in the hypothetical treatment than in the real treatment. Yet, in both cases, the coefficient of the hypothetical with cheap talk variable is insignificant at conventional levels, providing evidence that cheap talk might induce truthful responses among both dealers and nondealers. This evidence is consistent with the field data reported for the CEPA fundraising exercise and provides not only a robustness test of those results, but extends the findings to include experienced agents, who have not been affected by the cheap talk script in other valuation exercises (see, e.g., List, 2001; Lusk, 2003; Aadland and Caplan, 2003). We should exercise some caution in this case, however: using a one-sided alternative, the responses in the hypothetical with cheap talk treatment are significantly different at the $p < 0.11$ level.

Considering the second stage estimates, we first consider marginal valuation estimates for the attributes—centering and frayed corners. In all cases, a positive and statistically significant coefficient estimate suggests that agents in the real treatment are valuing centering and cornering when making valuation decisions. To examine hypothetical bias in the marginal value vector, we focus on the interaction terms. In every case we find that the hypothetical and hypothetical with cheap talk treatment variables interacted with these attributes are never statistically significant at conventional levels for either dealers or non-dealers. This result suggests that agents across the three treatments are not valuing the marginal attribute vector differently; suggesting the homogeneity null hypothesis (of no hypothetical bias) cannot be rejected for both the dealer and nondealer subsamples. This result provides tentative support for the use of hypothetical CE for intra-buy decisions.[16]

One potential explanation for the success of CE lies in the nature of the approach—it allows a subject to choose the most preferred alternative from a set of alternatives, which is a very common experience. For example, everyday tasks, such as walking the aisles of a department store, grocery store, or record shop, present very similar decision scenarios. In this respect, the decision process is much different from a typical CVM study, where subjects are presented with one hypothetical state of the world and must provide a maximum willingness to pay for the good. Rather, the CE approach encourages subjects to concentrate on the inherent trade-offs between attributes of the good instead of expressing a total value. This intuition is consistent with empirical evidence in Huber et al. (2002),

[16] Appendix E provides further empirical models that are meant to provide a robustness test of these results. We thank a reviewer for suggesting these models.

who find that experimental subjects view choice as the most realistic task as well as the mode in which they feel most confident making decisions. Some CVM studies have a dichotomous choice structure where they vote for or against an initiative. This is a common experience to individuals (and is similar to a "purchase" or "no purchase" situation in our choice experiment). However, by design, they do not allow individuals to trade-off between attributes and thus do not facilitate estimation of marginal values of attributes. Akin to the evidence from the psychology literature, which shows how decisions are determined partly by the ease of retrieval (e.g., the Availability Heuristic), this type of repeated decision procedure appears to diffuse the hypothetical bias found in other valuation institutions.

3.2 Preference Consistency

The above analysis provides some evidence that CE combined with cheap talk has the potential to yield reasonable estimates of economic values for non-marketed goods and services. Given that we have subjects making multiple decisions over highly similar choices, we are provided an opportunity to perform an *exploratory* test of validity: internal consistency of subject's preferences. Such an analysis permits an examination of the propensity of agents to make internally consistent decisions across the three treatments. In this casual analysis, we analyze violations of internal consistency that arise in one of two forms: *i*) when a subject declines to purchase a card superior to a card they consent to purchase in another question and *ii*) when a subject demonstrates a preference for one attribute over the other and then in a subsequent purchase decision reverses this preference. Peterson and Brown (1995) employ the paired comparison method to test for inconsistency in a manner analogous in certain respects to the technique we employ here.

The various proportions of subjects who have violations of consistency in each sub-sample of the data are reported in Table 5.[17] A first data pattern observed is that more experienced agents (dealers) tend to have fewer violations than their lesser experienced counterparts (non-dealers). This result also has some support statistically: upon pooling the data, a test of proportions suggests that fewer dealers commit violations compared to non-dealers at the $p < .05$ level. The result that a larger portion of inexperienced agents have violations of consistency has an interesting implication for CE: if the objective is to estimate preferences among those agents who minimally violate maintained economic assumptions, then individuals with greater levels of experience with the good are an important

[17] In sum, 11 subjects in our data set had more than one violation of consistency; of these 11, all had exactly two violations.

sample. Unfortunately, for the most important use of CVM, the valuation of non-market goods and services, such a sample is rarely available.

A second interesting pattern in the data is that across both dealers and non-dealers, the hypothetical with cheap talk treatment yields a larger proportion of violations. A test of proportions indicates that the proportion of subjects who committed at least one violation is significantly larger in the hypothetical with cheap talk treatment compared to the other two treatments at the $p < .05$ level.

Further insight can be gained by inquiring into the factors influencing this surprising finding. In the hypothetical with cheap talk treatment, 68.5 percent of subjects consent to purchase the Ryan card at least once and in the hypothetical treatment 67.6 percent do so. This difference is statistically negligible; however, when the total number of questions where subjects selected to purchase the card - recall that every subject responded to six questions - are compared we find that subjects purchased the card in 54.0% percent of questions in the hypothetical with cheap talk treatment, and in 63.6% of questions in the hypothetical treatment. This difference is significant at the $p < 0.05$ level. Moreover, in the hypothetical treatment, 30.9 percent of subjects selected to purchase the Ryan card for some questions and not for others, while in the hypothetical with cheap talk treatment this percentage was 55.3 percent.

If cheap talk works as intended (that is, if it induces subjects to reflect more on their true preferences and better consider their decisions) then these differences would be the result of subjects in the cheap talk treatment refusing to purchase the least desirable cards offered in the six questions. As is evidenced by the significant increase in the violations of consistency in the cheap talk treatment, however, subjects in the cheap talk treatment are not adjusting their behavior on the margin. Rather, subjects in the cheap talk treatment seem to be refusing purchase in an arbitrary manner: 33 of the 42 subjects who violate consistency in the cheap talk treatment do so by refusing purchase of a card superior to a card they consent to purchase in a prior or subsequent question.

These results can be interpreted as evidence that cheap talk may be effective in attenuating hypothetical bias because it introduces an additional bias into an individual's decision making that counteracts hypothetical bias. In our study cheap talk appears to bias subjects against purchasing, as subjects in the cheap talk treatment often select not to purchase the Ryan card in circumstances when doing so is inconsistent with their other decisions. This tentative conclusion certainly merits further research.[18]

[18] Cummings and Taylor (1999) discuss the possibility of cheap talk's effectiveness being the result of introducing new biases into subject's decision making, but do not find supporting evidence.

Table 5: Proportion of Subjects with Inconsistent Preferences

	Real	Hypothetical	Cheap Talk	Total:
Dealers (26/146)	0.125 (6/48)	0.143 (7/49)	0.265 (13/49)	0.178
Non-Dealers (65/225)	0.267 (20/75)	0.213 (16/75)	0.387 (29/75)	0.289
Total: (91/371)	0.211 (26/123)	0.185 (23/124)	0.339 (42/124)	0.245

Note: Data are taken from the second field experiment. Violations of internal consistency arise in one of two forms: i) when a subject declines to purchase a card superior to a card they consent to purchase in another question and ii) when a subject demonstrates a preference for one attribute over the other and then in a subsequent purchase decision reverses this preference. The proportions of subjects who have violations of consistency in each sub-sample of the data are reported in the Table. 11 subjects had more than one violation of consistency; of these 11 all had exactly two violations.

4. Conclusions

Contingent markets represent the only technique available to estimate the total value of a non-market good or service. To date, researchers have not been successful in designing an effective mechanism to eliminate hypothetical bias across all individual types in such markets, however. In this study, we examine individual behavior within choice experiments, which have an attractive theoretical structure, as the methodology conveniently combines Lancaster's (1966) characteristics theory of value with random utility theory (McFadden, 1974). While choice experiments are increasing in popularity, their demand-revealing properties in field applications are unknown.

This study examines the external validity properties of the choice approach by testing its performance in two field experiments. We find some evidence that the choice-based approach performs well for both private and public goods: in both the purchase and intra-buy decision hypothetical and real values are similar, and in all cases hypothetical choices in a cheap talk treatment are statistically indistinguishable from actual responses. This evidence suggests that the choice experimental approach *might* provide a valuable avenue to credibly estimate use and passive use values of non-market goods and services. Yet, a

caveat to this result is that cheap talk appears to induce internal inconsistency of subjects' preferences.

Appendix A.

Dear Ms. Doe,

As you are probably aware, Florida's recent rapid economic growth and development comes with potential environmental costs. Careful public policies are needed to protect local treasures such as the Everglades and the Florida panther while maintaining sustainable economic growth. To ensure that local decisions are made in the long-term interests of Florida citizens, we at the University of Central Florida are beginning a Center for Environmental Policy Analysis (CEPA). CEPA is a proposed research center to examine local, state, and global environmental issues such as air and water pollution, endangered species protection, and biodiversity enhancement. We believe that careful research will lead to solutions to important environmental problems.

CEPA will be housed in the Department of Economics in the College of Business. Although CEPA has some seed money available, we cannot begin operating until we have funded the equipment required for our researchers. Consequently, we are writing to ask for your help in creating CEPA at the University of Central Florida. You are part of a group of 1000 individuals to whom we are writing to fund this particular purchase. If we fail to raise the $3000 from this group of 1000 individuals, we will not be able to purchase the computer, but we will use the received funds to cover other operating expenditures of CEPA. If we do raise at least $3,000, we will purchase the computer and use any additional revenues above this threshold to fund CEPA's other needs. In either case, you will receive a note from us to let you know the status of your donation.

We would like you to consider making a contribution towards the purchase of a $3,000 computer to be used by researchers at CEPA by answering the choice question below.

Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
$20 donation	$5 donation	Status quo: no donation
UCF contributes $25	UCF contributes $5	UCF contributes $0

I hope you will join us in our commitment to sensible environmental policy by making a financial donation to CEPA. All donations are tax deductible. We have

enclosed a postage-paid return envelope for your convenience. For further information about CEPA, please see the enclosed brochure. Please contact me if you have any questions about this fundraising campaign. Thank you for your time.

Cheap Talk Script

In most questions of this kind, folks seem to have a hard time doing this. They act differently in a hypothetical situation, where they don't really have to pay money, than they do in a real situation, where they really have to pay money. We call this "hypothetical bias". "Hypothetical bias" is the difference that we continually see in the way people respond to hypothetical situations as compared to real situations. So, if I was in your shoes, and I was asked to make a choice, I would think about how I feel about spending my money this way. When I got ready to choose, I would ask myself: if this was a real situation, do I really want to spend my money this way?

Appendix B.

Welcome, and thanks for participating! Today you will have an opportunity to purchase the Nolan Ryan sportscard on the table.

Rules:

On the next page you will find six choice-based questions. I ask you to choose Column1, Column 2, or Column 3 for each question. After you have circled your choice for each question, I will roll a six-sided die (show them die) and whatever number is rolled will determine the question that is binding. Only one of your choices will be binding. For example, if I roll a 3, then question number 3 will be for real payment; the other questions will not be binding. Note that the die is "fair"—there is an equal probability that any of the six numbers will be rolled.

Do you have any questions?

Appendix C.

Before you make your choices I want to talk to you about a problem that we have in studies like this one. As I told you a minute ago, this is a hypothetical choice–not a real one. No one will actually pay money at the end. But, I also asked you to choose as though the result would involve a real cash payment. And that's the problem.

In most studies of this kind, folks seem to have a hard time doing this. They act differently in a hypothetical situation, where they don't really have to pay money, than they do in a real situation, where they really have to pay money. For example, in a recent study, several different groups of people bid in an auction. Payment was hypothetical for these groups, as it will be for you. No one had to pay money if they won the auction. The results of this study were that on average, across the groups, people overstated their actual willingness-to pay by 150 percent in the hypothetical auction. That's quite a difference, isn't it?

We call this "hypothetical bias". "Hypothetical bias" is the difference that we continually see in the way people respond to hypothetical situations as compared to real situations—just like the overbidding example presented above.

How can we get people to think about their choices in a hypothetical situation like they think in a real situation, where a person will really have to pay money? How do we get them to think about what it means to really dig into their pocket and pay money, if in fact they really aren't going to have to do it?

Let me tell you why I think that we continually see this hypothetical bias, why people behave differently in a hypothetical situation than they do when in a real situation. I think that when we behave in a hypothetical situation we place our best guess of what we would really like to do. But, when the choice is real, and we would actually have to spend our own money if we win, we think a different way: if I spend money on this, that's money I don't have to spend on other things ... we act in a way that takes into account the limited amount of money we have ... This is just my opinion, of course, but it's what I think may be going on in hypothetical situations.

So, if I was in your shoes, and I was asked to make several choices, I would think about how I feel about spending my money this way. When I got ready to choose, I would ask myself: if this was a real situation, and I had to pay $5, do I really want to spend my money this way?

Please keep this in mind when making your choices.

Appendix D.

1. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
80/20+FC	70/30+FC	Status quo: no purchase
at most 1 corner	at most 2 corners	
frayed	frayed	
$5 payment	$5 payment	

2. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3

75/25+FC at most 1 corner frayed $5 payment	70/30+FC at most 4 corners frayed $5 payment	Status quo: no purchase

3. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
80/20+FC at most 2 corners frayed $5 payment	70/30+FC at most 3 corners frayed $5 payment	Status quo: no purchase

4. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
60/40+FC at most 2 corners frayed $5 payment	70/30+FC at most 1 corner frayed $5 payment	Status quo: no purchase

5. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
60/40+FC at most 1 corner frayed $5 payment	80/20+FC at most 2 corners frayed $5 payment	Status quo: no purchase

6. Please circle your choice of Column 1, 2, or 3:

Column 1	Column 2	Column 3
60/40+FC at most 3 corners frayed $5 payment	70/30+FC at most 2 corners frayed $5 payment	Status quo: no purchase

Appendix E.

This Appendix contains further empirical results to examine the robustness of our results. First, Tables E1 and E2 run the exact same regressions as in the text but use only hypothetical and hypothetical with cheap talk dummies in the second stage and no interaction terms. Second, in Tables E3-E6, we run two sets of regressions with hypothetical and hypothetical with cheap talk dummies along with attribute variables in the first stage but with interaction terms in the second

stage, one set with a constant (Tables E3 and E4) and one set without a constant term (Tables E5 and E6).

Table E1: Sample Selection Model for Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"		
Regressors	**Coefficients**	**Z**
Hypothetical	0.67** (0.26)	2.6
Hypothetical with cheap talk	0.36 (0.23)	1.59
Question 1	-0.16 (0.07)	-2.28
Question 2	-0.20** (0.09)	-2.13
Question 3	-0.63** (0.11)	-5.87
Question 4	-0.26** (0.10)	-2.73
Question 5	-0.09 (0.08)	-1.14
Constant	0.66** (0.18)	3.70

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"		
Regressors	**Coefficients**	**Z**
Difference in centering	0.25** (0.03)	8.00
Difference in FC	0.25** (0.04)	6.81
Hypothetical	0.21 (0.16)	1.30
Hypothetical with cheap talk	0.10 (0.15)	0.68
Constant	-0.29 (0.11)	-2.60

Note:** denotes statistical significance at the $p < .05$ level.

Table E2: Sample Selection Model for Non-Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"		
Regressors	**Coefficients**	**Z**
Hypothetical	0.40** (0.19)	2.14
Hypothetical with cheap talk	0.12 (0.18)	0.67
Question 1	0.11 (0.06)	1.99
Question 2	0.06** (0.06)	0.97
Question 3	0.09** (0.07)	1.23
Question 4	-0.08** (0.07)	-1.14
Question 5	0.37 (0.06)	6.20
Constant	-0.52** (0.14)	-3.78

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"		
Regressors	**Coefficients**	**Z**
Difference in centering	0.15** (0.02)	6.21
Difference in FC	0.17** (0.03)	5.06
Hypothetical	-0.24 (0.16)	-1.46
Hypothetical with cheap talk	-0.07 (0.16)	-0.45
Constant	1.02 (0.12)	8.83

Note:** denotes statistical significance at the $p < .05$ level.

Table E3: Sample Selection Model for Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"		
Regressors	**Coefficients**	**Z**
Hypothetical	0.67** (0.26)	2.63
Hypothetical with cheap talk	0.36 (0.23)	1.61
Difference in centering	0.08 (0.02)	4.37
Difference in FC	0.01 (0.02)	0.42
Constant	0.43** (0.16)	2.65

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"		
Regressors	**Coefficients**	**Z**
Difference in centering	0.37** (0.07)	5.18
Difference in FC	0.43** (0.10)	4.40
Diff in centering × hypothetical with cheap talk	-0.03 (0.09)	-0.27
Diff in centering × hypothetical	-0.12 (0.08)	-1.45
Diff in FC × hypothetical with cheap talk	-0.14 (0.12)	-1.15
Diff in FC × hypothetical	-0.10 (0.11)	-0.94
Constant	-0.02 (0.17)	-0.13

Note:** denotes statistical significance at the $p < .05$ level.

Table E4: Sample Selection Model for Non- Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"

Regressors	Coefficients	Z
Hypothetical	0.39** (0.19)	2.11
Hypothetical with cheap talk	0.12 (0.18)	0.68
Difference in centering	0.05 (0.01)	5.42
Difference in FC	0.06** (0.01)	4.48
Constant	-0.47** (0.13)	-3.65

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"

Regressors	Coefficients	Z
Difference in centering	0.21** (0.08)	2.69
Difference in FC	0.40** (0.11)	3.75
Diff in centering × hypothetical with cheap talk	0.09 (0.09)	1.05
Diff in centering × hypothetical	-0.02 (0.08)	-0.24
Diff in FC × hypothetical with cheap talk	-0.18 (0.12)	-1.51
Diff in FC × hypothetical	-0.17 (0.12)	-1.45
Constant	0.36 (0.51)	0.71

Note:** denotes statistical significance at the $p < .05$ level.

Table E5: Sample Selection Model for Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"		
Regressors	**Coefficients**	**Z**
Hypothetical	0.68** (0.25)	2.68
Hypothetical with cheap talk	0.37 (0.23)	1.63
Difference in centering	0.08** (0.02)	4.38
Difference in FC	0.01 (0.02)	0.41
Constant	0.43** (0.16)	2.65

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"		
Regressors	**Coefficients**	**Z**
Difference in centering	0.37** (0.07)	5.54
Difference in FC	0.44** (0.09)	4.95
Diff in centering × hypothetical with cheap talk	-0.03 (0.09)	-0.29
Diff in centering × hypothetical	-0.12 (0.08)	-1.54
Diff in FC × hypothetical with cheap talk	-0.14 (0.12)	-1.24
Diff in FC × hypothetical	-0.11 (0.11)	-0.95

Note:** denotes statistical significance at the $p < .05$ level.

Table E6: Sample Selection Model for Non- Dealers

First Stage Estimation (Selection Equation): Dependent Variable is the Dummy "Purchased"		
Regressors	**Coefficients**	**Z**
Hypothetical	0.39** (0.19)	2.07
Hypothetical with cheap talk	0.11 (0.18)	0.63
Difference in centering	0.05** (0.01)	5.49
Difference in FC	0.06** (0.01)	4.49
Constant	-0.46** (0.13)	-3.58

Second Stage Estimation : Dependent Variable is the Dummy "Column 1"		
Regressors	**Coefficients**	**Z**
Difference in centering	0.23** (0.07)	3.42
Difference in FC	0.42** (0.09)	4.56
Diff in centering × hypothetical with cheap talk	0.10 (0.09)	1.09
Diff in centering × hypothetical	-0.01 (0.08)	-0.17
Diff in FC × hypothetical with cheap talk	-0.18 (0.12)	-1.46
Diff in FC × hypothetical	-0.15 (0.13)	-1.19

Note:** denotes statistical significance at the $p < .05$ level.

References

Aadland, D., and A. J. Caplan, "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias," *American Journal of Agricultural Economics*, 2003, 85(2): 492-502.

Adamowicz, Wiktor., Jordan Louviere, and Michael Williams, "Combining Revealed and Stated Preference Measures for Valuing Environmental Amenities," *Journal of Environmental Economics and Management*, 1994, 26(3): 271-292.

Adamowicz, Wiktor., Peter Boxall, Michael Williams, and Jordan Louviere, "Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation," *American Journal of Agricultural Economics*, 1998, 80(1): 64-75.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward Leamer, Roy Radner, and H. Schuman, "Report of the NOAA Panel on Contingent Valuation," *Federal Register,* 1993, 58(10): 4601-4614.

Bohm, P. "Estimating Demand for Public Goods: An Experiment," *European Economic Review*, 1972, 3(2):111-130.

Butler, J. S., and Robert Moffitt, "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," 1982, *Econometrica*, 50(3): 761-764.

Carlsson, Fredrik., and Peter Martinsson, "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" *Journal of Environmental Economics and Management*, 2001, 27(2): 179-192.

Carson, Richard T., Theodore Groves, and Mark J. Machina, "Incentive and Informational Properties of Preference Questions," University of California, San Diego, working paper, 2000.

Cummings, Ronald., Glenn Harrison, and Elisabet Rutström. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic Review*, 1995, 85(1): 260-266.

Cummings, Ronald., Elliott Steve, Glenn Harrison, and J. Murphy, "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, 1997, 105(3): 609-621.

Cummings, Ronald., and Laura Taylor, "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method," *American Economic Review*, 1999, 89(3): 649-65.

Diamond, Peter., and Jerry Hausman. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives*, 1994, 8(4): 45-64.

Foster, V., I. Bateman, and D. Harley, "Real and Hypothetical Willingness To Pay for Environmental Preservation: A Non-Experimental Comparison," *Journal of Agricultural Economics*, 1997, 48(2): 123-138.

Haab, Timothy C., Ju-Chin Huang, and John C. Whitehead. "Are Hypothetical Referenda Incentive Compatible? A Comment," *Journal of Political Economy*, 1999, 107(1): 186-196.

Hanley, Nick., Robert E Wright, and Wictor Adamowicz, "Using Choice Experiments to Value the Environment," *Environmental and Resource Economics*,1998, 11(3-4): 413-428.

Harrison, Glenn W. and List, John A., "Field Experiments," *Journal of Economic Literature*, December 2004, *42*(4), pp. 1009-55.

Huber, Joel., Dan Ariely, and Greg Fischer, "Expressing Preferences in a Principal-Agent Task: A Comparison of Choice, Rating, and Matching," *Organizational Behavior and Human Decision Processes*, 2002, 87(1): 66-90.

Lancaster, K. "A New Approach to Consumer Theory," *Journal of Political Economy*, 1974, 74(1): 132-157.

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

List, John A., and Craig Gallet, "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis," *Environmental and Resource Economics,* 2001, 20(3): 241-254.

List, John A., and David Lucking-Reiley, "The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign," *Journal of Political Economy* (2002), 110(1): 215-233.

List, John A., Robert Berrens, Alok Bohara, and Joe Kerkvliet, "Examining the Role of Social Isolation on Stated Preferences," *American Economic Review*, 2004, 4(3): 741-752.

List, John A. and Shogren, Jay. "Calibration of Willingness-to-Accept," *Journal of Environmental Economics and Management*, 2002, 43(2): 219-233.

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

Lusk, Jayson L. "Willingness-to-pay for Golden Rice," *American Journal of Agricultural Economics*, 2003, 85(4): 840-856.

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

Mixer, Joseph R. *Principles of Successful Fundraising: Useful Foundations for Successful Practice*, San Francisco: Jossey-Bass, 1993.

National Oceanic and Atmospheric Administration, "Natural Resource Damage Assessment: Proposed Rules," Federal Register, 4 May, 1994, 59: 23098-23111.

National Oceanic and Atmospheric Administration, "Natural resource damage assessments: Final rules," Federal Register, 5 January, 1996, 61:439.

Smith, V. Kerry. "Of Birds and Books: More on Hypothetical Referenda—Comment," *Journal of Political Economy*, 1999, 107(1):197-200.