Examining the Role of Social Isolation on Stated Preferences

By John A. List, Robert P. Berrens, Alok K. Bohara, and Joe Kerkvliet*

Benefit-cost analysis remains the central paradigm used throughout the public sector. A necessary condition underlying efficient benefit-cost analysis is an accurate estimate of the total value of the nonmarketed good or service in question. While economists have long measured the benefits of private goods routinely bought and sold in the marketplace, a much more difficult task faces the practitioner interested in estimating the total benefits of increased air and water quality, for example. In such cases, policy makers rely on stated preference methods (contingent markets) to provide signals of value. Recently there has been a lively debate about whether, and to what extent, "hypothetical bias" permeates benefit estimation in contingent markets.[1] This debate has proliferated among academics and practitioners over the past several decades, and continues to find its way into public disputes of damage assessment, development decisions, and discussions of optimal regulatory standards.

This study extends the debate in a new direction by taking advantage of a unique opportunity we were provided at the University of Central Florida (UCF), where we were approached to spearhead a capital campaign at UCF to fund a new Center for Environmental Policy Analysis (CEPA). The experimental design, which includes valuation decisions from nearly 300 subjects randomly placed into one of six treatment cells, permits an examination of the comparative static effects of varying social isolation while holding the other important facets of the valuation instrument constant. Our baseline treatments ask two different groups of respondents to vote Yes or No on contributing \$20 to provide start-up capital for CEPA (one treatment hypothetical and one treatment actual). In these two baseline treatments, similar to many practical methods of contingent valuation (CV) exercises that are carried out in practice (e.g., in-person, mail, or telephone), it is important to recognize that only the experimenter can observe each individual's response. In the third and fourth treatments, denoted Randomized Response, we again ask a hypothetical or actual question concerning a \$20 contribution, but we relax the degree of social pressure by using a randomized response format, which via delinking the observed and voting response ensures the subject that her stated preferences are unidentifiable.[2] These particular treatments resemble use of an anonymous ballot box approach to obtain individual values. Our final two treatments, labeled Peer Group, considerably decrease subject anonymity by randomly choosing 10 people to stand up and inform the group of their voting decision. These treatments bear resemblance to contingent surveys performed with small groups (or poorly controlled Web-based surveys).

The experimental results are interesting. Consonant with some previous studies, we observe signs of hypothetical bias. More importantly, we find that the difference between hypothetical and actual voting decisions is of roughly the same magnitude as the difference between actual voting decisions across treatments that vary

* List: AREC and Department of Economics, 2200 Symons Hall, University of Maryland, College Park, MD 20742, and National Bureau of Economic Research (e-mail: jlist@arec.umd.edu); Berrens and Bohara: Department of Economics, University of New Mexico, 1915 Roma NE, Albuquerque, NM 87131 (e-mail: rberrens@unm.edu; bohara@unm.edu); Kerkvliet: Department of Economics, 303 Ballard Extension Hall, Oregon State University, Corvallis, OR 97331 (e-mail: joe.kerkvliet@orst.edu). Three anonymous reviewers provided comments that improved the manuscript. Seminar participants at Cornell University, Harvard University, University of Arizona, University of Maryland, University of New Mexico, University of Pennsylvania, and the University of Colorado's Environmental Economics Workshop also provided useful suggestions. Richard Carson, Nick Flores, Glenn Harrison, Michael McKee, Kerry Smith, and Laura Taylor provided important insights on earlier drafts of this manuscript. We thank Apinya Thumaphipol for research assistance and Dean Thomas Keon for allowing us to use CEPA as a public good in our experiment. Any errors remain our own.

[1] We provide a more patient review of the various terms and the background of the debate in the next section.

[2] The randomized response approach to asking sensitive survey questions was introduced by Stanley Warner (1965).

social isolation. For example, across the three elicitation formats (Baseline, Randomized Response, and Peer Group) the largest percentage difference between Yes votes in the actual and hypothetical treatments is 14.5 percent; whereas the percentage differences across the actual voting decisions in the three formats are as large as 18 percent. This finding calls into question results from the plethora of validation studies that assume responses in the real payment treatment represent *true* preferences. We believe our results are fundamental to understanding the received evidence for mechanism design in CV, as they serve to highlight the notion that utilitarian elements and strategic reciprocity (e.g., from publicly advertising one's own goodwill) are confounded in interpreting signals of value, and that the potential biases introduced may be much larger than many expect.

I. Background, Experimental Design, and Hypotheses

One hallmark of public policy decision-making around the globe is a comparison of the costs and benefits of proposed policies. In the United States, President Clinton's Executive Order 12866, which reaffirmed the earlier executive order from the Reagan Administration, explicitly requires federal agencies to consider costs, benefits, and economic impacts of regulations prior to their implementation.[3] Given their flexibility to measure the monetary value of a wide variety of goods and services, CV methods have become a popular tool in practice for agencies to meet the Executive Order. CV refers to a set of survey-based approaches for eliciting Hicksian compensating or equivalent surplus values for a hypothetical change in a good or program. A contingent market (private or political) scenario is typically described for implementing a proposed change.

While the CV approach is practically quite important since it is literally the "only game in town" when it comes to measuring the total value of a nonmarketed good or service, critics contend that hypothetical bias severely limits its credibility (hypothetical bias is the difference between hypothetical and actual statements of value). In the burgeoning 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 (see, e.g., Ronald G. Cummings et al., 1995, 1997; Cummings and Laura O. Taylor, 1999; List, 2001).[4] Overall, scholars tend to find hypothetical bias in CV surveys (see List and Jason F. Shogren, 2002, for a review).

An interesting aspect of CV that has received considerably less attention is whether the survey administration mode is important—or likewise, whether social isolation affects stated preferences.[5] In practice, the mode of adminis-

[3] The more than 100 federal agencies issue approximately 4,500 new rulemaking notices each year. About 25 percent of those 4,500 are significant enough to warrant Office of Management and Budget (OMB) review. Of those, about 50–100 per year meet the necessary condition of being "economically significant" (more than $100 million in *either* yearly benefits or costs). Every economically significant proposal receives a formal analysis of the benefits/costs. The OMB establishes guidelines for the agencies on how to perform benefit-cost analysis.

[4] The interested reader should also see the exchange between Cummings et al. (1997), Timothy Haab et al. (1999), and V. Kerry Smith (1999).

[5] A related literature concerns the effect of decreasing anonymity or confidentiality in classic linear public goods experiments (see, e.g., David Masclet et al., 2003; Mari Rege and Kjetil Telle, 2004). For example, recent evidence shows that indirect or informal nonmonetary sanctions, conducted in some way through individuals' awareness of another subject's level of contribution, will increase cooperative behavior. While these studies help to underline the importance of communication between group members in efficiency-enhancing exercises, they do not speak to how social isolation between surveyor/agency and respondent influences stated preferences for nonmarketed goods and services. In this sense, contingent surveys could invoke much different preferences than those called upon in simple linear public goods: one may be a pure transfer and the other an efficiency-enhancing exercise. In addition, the extant linear public goods literature cannot provide a measure of the comparative-static effect of social isolation in the same framework as hypothetical bias is measured. Since policy makers have recognized hypothetical bias as an important shortcoming inherent in CV, in a practical sense an "apples-to-apples" measurement of the two effects is invaluable. Other fundamental differences between our research and the linear public goods literature, include, but are not limited to: (1) linear public goods games have not gathered data in an environment with complete confidentiality, as in our Randomized Response treatments; (2) linear public goods games deal with voluntary contributions, not with majority-rule voting on referenda with coercive taxes; (3) linear public goods games involve mechanisms that make no claims on incentive-compatibility (or demand revealing behavior), whereas there are arguments that referenda may be

tering contingent surveys has varied considerably. For example, Kenneth Arrow et al. (1993) advocate in-person interviews and several federal agencies currently use this approach (as well as mail or telephone surveys). Under this approach, it is important to recognize that respondents might be influenced by the presence of the surveyor, or the fact that their response is not entirely confidential. CV estimates have also been gathered via an anonymous ballot box approach, which various governmental agencies have utilized to gather preferences on a variety of issues. Likewise, at the other end of the "social isolation" spectrum, small groups and the Internet have both been used for CV surveys. The former approach provides a setting in which individual decisions are in every respect publicly observable and in the latter case the respondent may believe that the Internet is not 100 percent secure.

To extend the CV debate in a new direction, we attempt to mimic these various survey administration methods in various experimental treatments described below. As such, this research not only provides insights into the effects of the various survey administration methods, but also speaks to the plethora of validation studies that assume responses in the real payment treatment represent *true* preferences.

A. *Experimental Design*

Table 1 presents our 3 × 2 experimental design, and provides sample sizes for each treatment. Rows represent whether the treatment was hypothetical or actual and columns denote the three elicitation types (Peer Group, Baseline, and Randomized Response). Entries directly beneath the Baseline column represent value elicitation cells used in typical tests found in validation studies (List and Shogren, 2002). The other four cells are new to the literature and represent treatments that we use to disentangle the effect of social isolation on value. Each of the six treatments was carried out at UCF with students recruited from undergraduate courses in the College of Business. Each of the 268 students was provided a $20 participation fee.

TABLE 1—EXPERIMENTAL DESIGN

	Peer group	Baseline	Randomized response
Hypothetical	46	40	49
Actual	44	40	49

Note: Entries represent sample sizes. RR sample sizes are the summation of those who answered the real question and those who answered the innocuous question.

The public good valued concerns UCF's proposed Center for Environmental Policy Analysis (CEPA). All respondents are given the following information about the good and the referendum proposition:

> At this early stage, CEPA is a proposed research center to exam local and state environmental issues such as air and water pollution, endangered species protection, and biodiversity enhancement. Through careful research, solutions to important environmental problems can be advanced.
>
> The CEPA currently does not now have the funds required to begin. It will require $5,000 for start-up expenses. If everyone in this set of experiments were to contribute $20.00, these monies would be a sufficient beginning to cover the center's start-up cost.[6]
>
> We are going to have a vote to decide whether or not all of you will pay $20.00 for this purpose. We are voting on the following proposition:

incentive-compatible under certain conditions; (4) linear public goods games involve induced-value experiments rather than experiments with "home-grown" preferences; (5) linear public goods games typically examine data from an allocation of some proportion of an endowment rather than closed-ended, discrete voting behavior; (6) linear public goods games typically involve multiple-round experiments rather than single shot settings. The literature has shown that (2)–(6) can have important behavioral effects. Furthermore, a number of recent sources argue that classic public goods experiments with voluntary contributions provide a "particularly inappropriate market criterion to use as a base for assessing the validity of CV in field studies" (Gregory Poe et al., 2002, p. 106; Patricia Champ et al., 2002).

[6] The astute reader will note that required start-up monies are larger than what could be gathered in these treatments. We were careful to note to subjects: "If everyone in this set of experiments were to contribute $20.00, these monies would be a sufficient beginning to cover the center's start-up cost." In this study we report results from six treatments. The set of experiments included several different treatments examining economic theory.

Everyone in the room will contribute $20.00 to the UCF Center for Environmental Policy Analysis. The contribution will be used for the purpose of covering start-up expenses.[7]

Before the vote, participants were informed exactly how the majority rule referendum format operates, how monies will be collected (in the actual referenda cases), and given the opportunity for questions on the referendum mechanism. All respondents were then presented with a referendum-voting question with a fixed $20 payment amount for the same public good.[8]

As summarized in Table 1, the elicitation types are denoted *Baseline,* which is consistent with typical experimental treatments and many CV administration methods (in-person, mail, and telephone interviews) in that only the surveyor/government agency can observe each subject's response. In this case, the experimenter is aware of the choices since subjects' votes can be linked with a survey sheet that contains their name and other individual-specific particulars (see Appendix B). The second elicitation type is the *Randomized Response* (RR hereafter) approach, where the interviewer knows the individual response (Yes or No), but does not know whether it was to the referendum question, or to an alternative innocuous question.[9] Accordingly, we expect that any upward response effects associated with reciprocity and positive status effects are eliminated in the RR since the individual voting responses are entirely confidential. The final treatment is labeled *Peer Group,* and is identical to the *Baseline* treatment, except in this case 10 people are chosen randomly to stand up and inform the group of their response. Thus, in this treatment each individual answers the referendum question knowing that there is a possibility of having his or her response made public.[10] Relative to the baseline, the RR treatment provides an increased degree of social isolation, while the Peer Group treatment provides a decreased degree of social isolation. Crossing each of these elicitation techniques with a hypothetical and actual treatment produces the six treatments in Table 1: BASE-HYPO, BASE-ACTUAL, RR-HYPO, RR-ACTUAL, PEER-HYPO, and PEER-ACTUAL.[11]

Before proceeding to a discussion of the hypotheses, it is worthwhile to explain briefly our application of the RR method since it is rarely used in economics. We use the "unrelated question" design of Bernard G. Greenberg et al. (1969), which is a randomization process that directs the respondent, with a probability con-

[7] To ensure subjects that the money would actually go to CEPA, we noted in the experimental instructions: "We will not send cash. I will take your cash, write this check (show check) for ($n \times$ \$20.00) and the check will be mailed to the center. I will put the check in this stamped envelope (show envelope) addressed to the center. I will ask one of you to put the envelope in the mailbox downstairs. When I receive a receipt for the money from the center, I will make it available for your inspection in front of room 319 in the CBA building."

[8] This design choice closely follows Cummings et al. (1997), among others, who provide each subject with a $10 participation fee and ask subjects to vote on a proposal to give $10 for the production and distribution of citizens' guides in New Mexico. We also ran pilot treatments where subjects first earned at least $20. These results, which are available upon request, provide qualitatively identical insights to the data described herein. One difference is that the percentage of Yes votes is lower in each treatment.

[9] This approach was used to mimic the anonymous ballot box method. In very small scale pilot tests the two approaches yielded qualitatively similar insights so we opted to use the RR approach because it allows us to match subject-specific characteristics with responses. This area is ripe for future research.

[10] In the *Peer Group* treatment, directly after the example of how the referendum works, and before the actual referendum question, participants are told on the written questionnaire: "Please note that we will be calling on 10 people to stand up and announce their response. For example, if I call John Doe, he must stand up and announce to the class how he voted." Subjects were aware that the experimental monitor had the survey and voting sheets in hand when he called the subjects' names to announce their vote. Thus, the subjects must reveal truthfully (in practice all did so). Note, also, that the referendum worked exactly as in the other treatments—it was based on voting from the entire group, not merely the 10 subjects who were called to announce their vote. This approach is intended to represent contingent surveys carried out with small groups and poorly controlled Web-based CV surveys.

[11] Appendix A contains a copy of the Baseline actual treatment instructions. Instructions for the other treatments were identical except for the necessary changes. We interchange "hypothetical" and "HYPO" hereafter. At this point, we should note that subjects were given the chance to leave the experiment after they understood the rules. No subjects exited early. And, an astute reader will recognize that there are slight differences in group size across treatments (see Table 1), which could influence voting behavior because of its effect on the likelihood that someone is pivotal. To ensure that group sizes were not a confounding factor, we had subjects enter the large room from the rear and used cardboard row dividers to ensure subjects could not determine the exact group size.

trolled by the analyst, to give a Yes or No answer to either an unrelated question or the sensitive question (e.g., the CEPA voting question). The text for our RR treatment included:

> So that only you will know which question you answered, it will first be necessary that you compute a random number from your Social Security Number. **THIS RANDOM NUMBER IS NOT YOUR SOCIAL SECURITY NUMBER AND CANNOT BE USED TO FIND YOUR SOCIAL SECURITY NUMBER.** This random number is the sum of the last four digits of your Social Security Number. For example, if your Social Security Number is:
>
> 517-48-1234: $1 + 2 + 3 + 4 = 10$
>
> Now compute your random number using your Social Security Number. Do **NOT** write or speak this number, but remember it for the first question. Your answer cannot be traced to you, nor do we have any interest in doing so.
>
> 1. This question requires a Yes or No answer to one of the following two questions. If your random number is between 0 and 10, answer question A below. If your random number is between 11 and 36, answer question B below.

Where the innocuous question A was whether the respondent's mother's birthday is in a particular month or set of months (May and June), and question B was the referendum.

B. *Hypotheses*

The experimental design permits us to test a number of important hypotheses related to the hypothetical bias debate within the CV literature. First, we can pool the data and examine if actual and hypothetical responses are in accord:

$$\boldsymbol{H}_1\text{: Prob(Yes)}_{\text{HYPO}} \neq \text{Prob(Yes)}_{\text{ACTUAL}}.$$

Second, we can disaggregate the data by controlling for elicitation type, and test for differences between hypothetical and actual responses within each social isolation regime:

$$\boldsymbol{H}_2\text{: Prob(Yes)}_{\text{RR-HYPO}} \neq \text{Prob(Yes)}_{\text{RR-ACTUAL}},$$

$$\boldsymbol{H}_3\text{: Prob(Yes)}_{\text{BASE-HYPO}} \neq \text{Prob(Yes)}_{\text{BASE-ACTUAL}},$$

$$\boldsymbol{H}_4\text{: Prob(Yes)}_{\text{PEER-HYPO}} \neq \text{Prob(Yes)}_{\text{PEER-ACTUAL}},$$

In addition to these tests of hypothetical versus actual voting decisions, our experimental design also permits an examination of social isolation effects via six distinct tests of the following spirit:

$$\boldsymbol{H}_5\text{: Prob(Yes)}_{ik} \neq \text{Prob(Yes)}_{jk},$$

where $i \neq j$ refers to the survey regime and k denotes hypothetical or actual. Concerning expectations relative to the baseline treatment, our working hypothesis for these tests is that the RR approach will reduce the probability of a Yes vote, and that overt social pressure (Peer Group) will increase the probability of a Yes vote.[12]

II. Experimental Results

Table 2 presents descriptive statistics and a selected set of respondent characteristics for each treatment. A first important result displayed in columns 4 and 8 of Table 2 is that

[12] In addition to unconditional tests, we examine our hypotheses within a conditional setting—e.g., estimate a probit regression of the form $\Pr(\text{Yes}) = f(X'\boldsymbol{\lambda})$, where X is a set of explanatory variables gathered from the survey in Appendix B, and $\boldsymbol{\lambda}$ is the vector of associated coefficients. We use the sequential testing procedure originally proposed in Joffre Swait and Jordan Louviere (1993). First, a pooled probit model is estimated with the optimum variance ratio (σ_1/σ_2) for any two samples (1 and 2) derived from a grid search, where, for example, the standard deviation of the control sample, σ_1, is set equal to one, and the standard deviation of the treatment sample is estimated freely. The null of the equality of the coefficient vector is tested by comparing this restricted log-likelihood value to the unrestricted log-likelihood value (sum of the two separate probits). An insignificant χ^2 value indicates that the coefficient vectors are not statistically different. The difference in the two variances is tested by comparing the unrestricted log-likelihood value with the restricted likelihood (a simple pooled probit). Finally, in order to conduct these hypotheses tests for the RR treatment, estimating the probit model for the RR sample must take account of the known randomization mechanism. Our approach follows that of Berrens et al. (1997). The primary caveat to the notation of Berrens et al. (1997: pp. 255–58) is that our "bid" (t) was fixed at \$20. Our calculated value for the probability that the constructed random number was between 0–10 ("$1-\rho$") was 0.077, as constructed from the mean of a matching random sample of 300 UCF students to our experimental sample (Berrens et al., 1997, used a value of 0.08). We also assumed that the probability a mother's birthday occurred in a given month ("$\hat{p}$") was 0.083.

TABLE 2—DESCRIPTIVE STATISTICS, SPLIT SAMPLES MEANS FOR SELECTED VARIABLES

Variable	HYPO-BASE [n = 40]	HYPO-RR [n = 49]	HYPO-PEER [n = 46]	HYPO TOTAL [n = 135]	ACTUAL-BASE [n = 40]	ACTUAL-RR [n = 49]	ACTUAL-PEER [n = 44]	ACTUAL TOTAL [n = 133]	TOTAL [n = 268]
Pr(YES)	0.525	0.330	0.630	0.489	0.380	0.200	0.500	0.353	0.422
	(0.506)	(0.481)	(0.488)	(0.502)	(0.490)	(0.435)	(0.506)	(0.484)	(0.496)
LnINC	0.154	0.238	0.073	0.157	0.032	0.060	0.240	0.111	0.134
	(0.730)	(0.740)	(0.751)	(0.738)	(0.807)	(0.805)	(0.810)	(0.806)	(0.77)
MALE	0.550	0.592	0.522	0.556	0.450	0.531	0.568	0.519	0.537
	(0.504)	(0.497)	(0.505)	(0.499)	(0.504)	(0.504)	(0.501)	(0.502)	(0.500)
LnAGE	−1.488	−1.481	−1.519	−1.496	−1.513	−1.512	−1.457	−1.494	−1.495
	(0.160)	(0.184)	(0.121)	(0.158)	(0.159)	(0.123)	(0.172)	(0.153)	(0.155)
URBAN	0.725	0.510	0.609	0.607	0.700	0.571	0.546	0.602	0.605
	(0.452)	(0.505)	(0.493)	(0.490)	(0.464)	(0.500)	(0.504)	(0.491)	(0.490)
ENVM	0.075	0.041	0.065	0.059	0.050	0.061	0.091	0.068	0.063
	(0.267)	(0.200)	(0.250)	(0.237)	(0.221)	(0.242)	(0.291)	(0.252)	(0.244)

Notes: Numbers in parentheses are standard errors. Pr(Yes) is the probability of a Yes response to the offered \$20 payment. LnINC is the natural log of the INC variable; MALE is a dummy indicator variable for gender, with 1 = male and 0 = female; LnAGE is the natural log of the respondent's (age in years/100); URBAN is a dummy indicator variable, with 1 = from an urban area, and 0 = otherwise; ENVM is a dummy indicator variable, with 1 = member of environmental group, and 0 = otherwise. For RR subsamples, we present the adjusted Pr(Yes) (see James Alan Fox and Paul E. Tracy, 1988).

upon pooling the data, an initial tendency observed is that the hypothetical treatments garner more votes in the affirmative than the actual regimes (48.9 percent versus 35.3 percent). This result is supported statistically, as various tests provide evidence in support of alternative hypothesis $\boldsymbol{H}_1$: a significant difference exists between hypothetical and actual responses.[13] Upon disaggregating the data into elicitation types, we find that for all possible binary comparisons the proportion of Yes votes is greater in the hypothetical treatment compared to the actual treatment, which is consistent with empirical results observed in some other experimental settings (e.g., Cummings et al., 1995, 1997; List, 2001; Poe et al., 2002). For example, whereas 52.5 percent of individuals voted Yes in the hypothetical baseline treatment, only 38 percent voted Yes in the actual baseline. In sum, the discrepancy between actual and hypothetical statements is consistent across elicitation type, and ranges from 13 percent to 14.5 percent.[14]

In Table 3, we examine hypothetical bias more closely by presenting results from a variety of tests using binary comparisons of all treatment types. For completeness, we present results from three different sets of tests: contingency table Pearson χ^2, population proportions test of independent binomial experiments, and Fisher's Exact test. Findings suggest that individually there is not strong evidence of hypothetical bias: results in the first six rows of Table 3 imply that within each elicitation type we should not reject the null hypothesis of similar hypothetical and actual responses. Yet when we apply a more powerful aggregate test statistic (Kelly Busche and Peter Kennedy, 1984) that takes into account the fact that the underlying proportion of Yes votes may vary across treatments, even though they may still be equivalent (under the null hypothesis), we find evidence that supports rejection of the null hypothesis.[15]

[13] The statistical test results for hypothetical versus actual (null hypothesis of no difference) are: Pearson χ^2 value = 5.02** (p-value = 0.02); one-tailed Fisher's test p-value = 0.02**; two-tailed Fisher's test p-value = 0.03**; population proportions test $|z|$-statistic = 2.35**; where: ** indicates significant at the $p < 0.05$ level.

[14] This constancy of hypothetical bias across the three referenda is interesting in its own right. While we leave the pursuit of this puzzle for further research, we speculate that it may represent an autonomic first response in this environment. Under this interpretation, this insight is consonant with theories of experiential learning (e.g., Elana Michelson, 1999): absent experience with the good and elicitation institution, hypothetical upward bias is an autonomical first response.

[15] This test relies on independence and the fact that the statistic for each treatment has an approximately normal distribution with mean 0 and variance 1. Recall that the sum of n independent, normally distributed random variables is

TABLE 3—DESCRIPTIVE STATISTICS, TEST RESULTS FOR SPLIT SAMPLE COMPARISONS

Sample	Treatment split samples tested [n]	Proportion of Yes responses	χ^2 statistic (p-value)	Pairwise comparison of proportions $\lvert z\rvert$-statistic	Fisher's test p-value (2 and 1 tailed)	Comment
BASE	HYPO [40]	0.53	1.82 (0.18)	1.36	0.26 0.13	No Difference HYPO vs. ACTUAL
	ACTUAL [40]	0.38				
PEER	HYPO [46]	0.63	1.56 (0.21)	1.25	0.29 0.15	No Difference HYPO vs. ACTUAL
	ACTUAL [44]	0.50				
RR	HYPO [49]	0.33	1.88 (0.17)	1.47	0.25 0.12	No Difference HYPO vs. ACTUAL
	ACTUAL [49]	0.20				
ACTUAL	BASE [40]	0.38	1.33 (0.24)	1.12	0.28 0.18	No Difference BASE vs. PEER
	PEER [44]	0.50				
ACTUAL	BASE [40]	0.38	3.18* (0.07)	1.65*	0.10* 0.06*	Significant Difference BASE vs. RR
	RR [49]	0.20				
ACTUAL	PEER [44]	0.50	8.99*** (0.00)	3.17***	0.004*** 0.003***	Significant Difference PEER vs. RR
	RR [49]	0.20				
HYPO	BASE [40]	0.53	0.62 (0.43)	0.94	0.55 0.28	No Difference BASE vs. PEER
	PEER [46]	0.63				
HYPO	BASE [40]	0.53	3.57* (0.06)	1.93*	0.08* 0.05**	Significant Difference BASE vs. RR
	RR [49]	0.33				
HYPO	PEER [46]	0.63	8.79*** (0.00)	3.07***	0.004*** 0.003***	Significant Difference PEER vs. RR
	RR [49]	0.33				

Notes: *, **, *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively. In calculating the proportion of Yes responses for RR subsamples, the observed count of Yes responses has been converted to an adjusted estimate of Yes responses, using the formula of Fox and Tracy (1988).

Interestingly, empirical evidence in Table 3 suggests that the proportion of Yes votes in the RR treatments is significantly *lower* than the proportion of Yes votes in comparable BASE and PEER treatments. Indeed, in terms of statistical and economic significance, the data suggest that the effect of varying social isolation is in the range of the comparative-static effect of varying the monetary consequentiality of the decision. For example, a comparison between the percentage of Yes votes across BASE and RR (PEER) yields a difference of 0.18 (0.12) in

a normally distributed random variable with mean (variance) equal to the sum of the means (variances). In our case the statistic equals the sum of the three independent statistics divided by the square root of three.

TABLE 4—PROBIT MODEL RESULTS

Variable	BASE sample [n = 80]	PEER sample [n = 90]	RR sample [n = 98]	Full sample [n = 268]	Full sample [n = 268]
CONSTANT	−0.44*	0.02	−0.82**	−0.78***	−0.74***
	(−1.93)	(0.07)	(−2.63)	(−4.08)	(−3.25)
LnINC	0.37**	0.28	1.11**	0.46***	0.46***
	(1.97)	(1.56)	(2.57)	(4.00)	(4.00)
MALE	0.21	−0.15	−0.43	−0.11	−0.11
	(0.72)	(−0.55)	(−1.22)	(−0.69)	(−0.67)
HYPO	0.34	0.38	0.26	0.36**	0.28
	(1.17)	(1.42)	(0.76)	(2.19)	(0.96)
BASE				0.47**	0.42
				(2.29)	(1.45)
PEER				0.77***	0.69**
				(3.87)	(2.39)
HYPO*BASE					0.08
					(0.21)
HYPO*PEER					0.15
					(0.38)
Log-likelihood value	−52.08	−59.50	−49.30	−164.85	−164.78
AIC	112.16	127.008	106.589	341.69	345.56
Maddala R^2	0.072	0.054	0.181	0.122	0.123

Notes: Numbers in parentheses are *t*-statistics; *, **, *** denote significance at the 0.10, 0.05, and 0.01 levels, respectively. The RR sample uses the DC-CV modeling approach detailed in Berrens et al. (1997).

the actual treatment and 0.20 (0.10) in the hypothetical treatment; overall these differences are statistically significant at conventional levels (except for the BASE versus PEER comparisons, which are nearly significant using a one-sided alternative). If one considers the extreme comparison case—RR versus PEER—we find that whereas 50 percent of respondents voted Yes in the actual PEER treatment, only 20 percent voted Yes in the RR actual treatment, a difference that is statistically significant at the $p < 0.01$ level.[16]

While these results are certainly suggestive that social isolation matters a great deal, no attempt has been made to control for the possible influence of socioeconomic characteristics, and for this we turn to empirical estimates from various probit regressions. Table 4 presents empirical estimates from five distinct probit models, which each control for income (LnINC) and gender (MALE).[17] The empirical models differ by the sample used and the treatment dummy variables included. Results reported in columns 1–3 are from separate elicitation types and include a hypothetical dichotomous variable (HYPO = 1 if hypothetical referendum, 0 if actual referendum). In each case the estimated coefficient on this dummy variable is positive, but not statistically different from zero. This finding is in line with the individual results reported above.

The model reported in column 4 uses the pooled sample, and in addition to the hypothetical dummy variable it controls for the elicitation type via inclusion of dichotomous variables indicating Baseline and Peer Group treatments (BASE and PEER). In these cases, the estimated coefficient on hypothetical treatment is positive

[16] Our finding that varying social isolation has a similar percentage effect on voting patterns as varying the monetary consequentiality of the decision may even be considered a lower bound estimate since the difference between hypothetical and actual voting patterns may be increased by our use of an open referendum (Richard Carson et al., 1999). Further, the influence of social isolation may be muted if subjects believed that the experimenter had access to their social security numbers.

[17] In the specifications of Table 4, we control for income and gender (LnINC and MALE). The income variable is significant in a number of models. While never significant in our setting, gender has been shown to be a significant explanatory variable in a number of experimental studies (see, e.g., James Andreoni and Lise Vesterlund, 2001). A wide variety of alternative specifications were also evaluated—we provide these estimates in tabular form on the AER Web site: http://www.aeaweb.org/aer/contents/. The qualitative conclusions on the treatment variables remain unchanged.

and significant at the $p < 0.05$ level. In addition, the estimated coefficients on BASE and PEER are both positive and significant at the $p < 0.05$ and $p < 0.01$ levels. Hence, in this particular model there is evidence for both hypothetical bias and a social isolation effect, and they are of roughly similar magnitudes, though the social isolation effect is slightly larger.[18]

One criticism of our dummy variable approach is that it forces the variances of the different treatment samples to be equivalent (Haab et al., 1999). To amend this potential shortcoming, we ran separate probit models with identical sets of explanatory variables (CONSTANT, LnINC, and MALE) and no treatment dummy variables. Summary results, which are available on the AER Web site (http://www.aeaweb.org/acr/contents/) in tabular form, reveal evidence consonant with Haab et al. (1999) in that there are differences in variances within our data: in comparing HYPO versus ACTUAL, the evidence supports the hypothesis of a significant difference at the $p < 0.01$ level. The null hypothesis between the HYPO and ACTUAL variances, however, cannot be rejected for the RR and the PEER treatments. Further, when looking at the HYPO and ACTUAL samples separately, there is evidence of a treatment effect, where the variance in the RR treatment is significantly different from that of the BASE and PEER treatments. These results confirm our unconditional findings reported above, and suggest the significant role that social isolation plays in the valuation process.

[18] The importance of social isolation is further illustrated in column 5 of Table 4—a pooled model that includes interaction terms HYPO*BASE and HYPO*PEER. In this particular specification, estimated coefficients on HYPO, HYPO*BASE, BASE, and HYPO*ACTUAL are not significantly different from zero, either individually or jointly. Yet the treatment variables remain statistically significant (jointly), and we again find that subjects randomly inserted into the PEER treatment vote Yes significantly more than respondents placed in the RR treatment.

III. Discussion

Recently there has been a lively debate about whether, and to what extent, "hypothetical bias" permeates benefit estimation in contingent markets. Given that benefit-cost analyses are required at the federal level, and increasingly at the state level, investigating potential biases in contingent valuation has great practical importance. This paper extends the debate in a new direction by exploring the link between social isolation and stated preferences. Examining data from nearly 300 subjects placed randomly into one of six experimental treatment cells, we find that social isolation plays a considerable role. Indeed, its magnitude is roughly comparable to the degree of hypothetical bias observed.

Besides its importance for practical implementation of contingent valuation, our findings raise serious concerns about the experimental results in the literature purporting to measure hypothetical bias given the specific social context in which some of the studies have been conducted. For example, are "actual" statements of value in these experiments providing accurate signals of *true* preferences? And, what is the correct benchmark if the degree of social isolation is not controlled? While our study only pertains directly to the referendum voting institution, our findings raise the specter that social isolation effects may be important in every elicitation format, including open-ended valuation questions, choice experiments, dichotomous choice questions, etc. This effect may be especially important when the survey mode involves direct social interaction, as in personal interviews, or when the issue is salient, as is the case with many environmental matters. It seems clear these settings may well induce respondents to include any number of utility-enhancing values that come from publicly advertising one's own goodwill. But, since these "externality-type" values are not germane to the good in question, rather to a class of goods, it is incorrect to lump them with any particular good's value.

Appendix A: Instructions for Actual Baseline

The following questionnaire concerns the University of Central Florida's proposed Center for Environmental Policy Analysis (CEPA). It should take only several minutes to complete; all answers will be treated as confidential information and are an important input to our study. Except when asked for a specific number, most of the questions can be answered simply by checking the appropriate response. Your time and consideration are appreciated.

At this early stage, CEPA is a *proposed* research center to examine local and state environmental issues such as air and water pollution, endangered species protection, and biodiversity enhancement. Through careful research, solutions to important environmental problems can be advanced.

The CEPA currently does not now have the funds required to begin. It will require $5,000 for start-up expenses. If everyone in this set of experiments were to contribute $20.00, these monies would be a sufficient beginning to cover the center's start-up cost.

We are going to have a vote to decide whether or not all of us will pay $20.00 for this purpose. We are voting on the following proposition:

Everyone in the room will contribute $20.00 to the UCF Center for Environmental Policy Analysis. The contribution will be used for the purpose of covering start-up expenses.

Here's how the vote works.

1. If more than 50% of you vote "yes" on this proposition, all of you will pay $20.00—I will collect $20.00 from each of you—and we will send this money to the UCF CEPA with instructions that the money is to be used to fund start-up expenses.
   We will not send cash. I will take your cash, write this check (show check) for ($n \times$ $20.00) and the check will be mailed to the center. I will put the check in this stamped envelope (show envelope) addressed to the center. I will ask one of you to put the envelope in the mailbox downstairs. When I receive a receipt for the money from the center, I will make it available for your inspection in front of room 319 in the CBA building.
2. If 50% or fewer of you vote "yes" on this proposition, no one will pay $20.00, we will not send a check to the center and the start-up expenses will not be gathered.
   We are now passing out a ballot. Remember how the vote works. If more than 50% vote "yes" we will collect $20.00 from each of you, and we will mail this check to the center right here today. If 50% or less vote "yes," no one will pay $20.00, and we will not mail this check to the center. Any questions?

Appendix B: Survey Summary

These answers are strictly confidential and are needed for statistical purposes only.
NAME: __________
Q.2. AGE: ____
Q.3. Gender: Male Female
Q.4a. Have you decided on a major field of study: ____ Yes
____ No
Q.4b. If Yes, in what academic college at UCF is your major field?
____ Agricultural ____ Business ____ Pre-engineering ____ Engineering ____ Forestry
____ Home Economics ____ Science ____ Pharmacy ____ Health and Human Performance
____ Special Graduate Programs ____ Veterinary Medicine ____ Health and Public Affairs
____ Arts and Sciences (Sciences) ____ Arts and Sciences (Arts)
Q.5. Which of the following *best* describes your background: ____ Rural ____ Urban
Q.6. Approximate household income per month:

____ 0–$300	____ $301–$500
____ $501–$800	____ $801–$1000
____ $1001–$1500	____ $1501–$2000
____ $2001–$2500	____ $2501+

Q.7. What is your response to the following statement: Protecting the environment is so important that requirements and standards cannot be too high, and continuing environmental improvements must be made regardless of cost.

____ Strongly agree ____ Agree ____ Disagree ____ Strongly disagree

Q.8. Are you a member of any local, state, or national environmental organizations such as the Sierra Club, Nature Conservancy, etc.?

____ Yes ____ No

Q.9. Marital status: ____ Married ____ Not married

Q.10. Region in which you grew up:

____ Northeast ____ South
____ Midwest ____ Mountain States
____ Pacific Coast

REFERENCES

Andreoni, James, and Vesterlund, Lise. "Which Is the Fair Sex? Gender Differences in Altruism." *Quarterly Journal of Economics,* February 2001, *116*(1), pp. 293–312.

Arrow, Kenneth; Solow, Robert; Portney, Paul R.; Leamer, Edward E.; Radner, Roy and Schuman, Howard. "Report of the NOAA Panel on Contingent Valuation." *Federal Register,* January 1993, *58*(10), pp. 4602–14.

Berrens, Robert P.; Bohara, Alok K. and Kerkvliet, Joe. "A Randomized Response Approach to Dichotomous Choice CV." *American Journal of Agricultural Economics,* February 1997, *79*(1), pp. 252–66.

Busche, Kelly and Kennedy, Peter. "On Economists' Belief in the Law of Small Numbers." *Economic Inquiry,* October 1984, *22*(4), pp. 602–03.

Carson, Richard T.; Groves, Theodore and Machina, Mark. "Incentive and Informational Properties of Preference Questions." Working paper, University of California, San Diego, 1999.

Champ, Patricia A.; Flores, Nicholas E.; Brown, Thomas C. and Chivers, James. "Contingent Valuation and Incentives." *Land Economics,* November 2002, *78*(4), pp. 591–604.

Cummings, Ronald G.; Elliott, Steven; Harrison, Glenn W. and Murphy, James. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy,* June 1997, *105*(3), pp. 609–21.

Cummings, Ronald G.; Harrison, Glenn W. and Rutstrom, E. Elisabet. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review,* March 1995, *85*(1), pp. 260–66.

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

Fox, James Alan and Paul E. Tracy. *Randomized response: A method for sensitive surveys.* Beverly Hills, CA: Sage Publications, 1986.

Greenberg, Bernard G.; Abul-Ela, Abdel-Latif A.; Simmons, Walt R. and Horvitz, Daniel G. "The Unrelated Question Randomized Response Model: Theoretical Framework." *Journal of the American Statistical Association,* June 1969, *64*(326), pp. 520–39.

Haab, Timothy C.; Huang, Ju-Chin and Whitehead, John C. "Are Hypothetical Referenda Incentive Compatible? A Comment." *Journal of Political Economy,* February 1999, *107*(1), pp. 186–96.

List, John A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review,* December 2001, *91*(5), pp. 1498–507.

List, John A. and Shogren, Jason F. "Calibration of Willingness-to-Accept." *Journal of Environmental Economics and Management,* March 2002, *43*(2), pp. 219–33.

Masclet, David; Noussair, Charles; Tucker, Steven and Villeval, Marie-Claire. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review,* March 2003, *93*(1), pp. 366–80.

Michelson, Elana. "Carnival, Paranoia, and Experiential Learning." *Studies in the Education of Adults,* October 1999, *31*(2), pp. 141–63.

Poe, Gregory L.; Clark, Jeremy E.; Rondeau, Daniel and Schulze, William D. "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics,* September 2002, *23*(1), pp. 105–31.

Rege, Mari and Telle, Kjetil. "The Impact of Social Approval and Framing on Cooperation in Public Good Situations." *Journal of Public Economics,* July 2004, *88*(7–8), pp. 1625–44.

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

Swait, Joffre and Louviere, Jordan. "The Role of the Scale Parameter in the Estimation and Comparison of Multinational Logit Models." *Journal of Marketing Research,* August 1993, *30*(3), pp. 305–14.

Warner, Stanley. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association,* March 1965, *60*(309), pp. 63–69.