The use of a real-money experiment in a stated-preference survey

by

John Horowitz
Department of Agricultural and Resource Economics
University of Maryland
Horowitz@arec.umd.edu

November 7, 2005

THE USE OF A REAL-MONEY EXPERIMENT IN A STATED-PREFERENCE SURVEY[1]

1. INTRODUCTION

Environmental economics divides into two sub-disciplines: valuation and regulation. Regulation is the analysis of how regulations work, how much they cost, and how they can be improved. Valuation is the analysis and measurement of the benefits of environmental quality. It primarily means assigning a dollar value to environmental quality. Valuation refers to both revealed and stated preference techniques. Because there are many environmental scenarios for which revealed preference data are difficult or impossible to obtain, a large portion of environmental valuation relies on stated-preference techniques. Stated preference, a general category of which contingent valuation is a part, means the use of surveys and similar instruments to elicit a value for environmental quality.

Environmental valuation is an essential component of economics but its results have been greeted with indifference, skepticism, suspicion, or even hostility.[2] This skepticism extends to both revealed and stated preference findings, although different concerns underlie each of these techniques. This chapter examines stated preference techniques, which I sometimes refer to simply as "valuation surveys."

I focus on one of the key reasons for skepticism about stated preference techniques, namely, their hypothetical nature. Because stated preference is hypothetical by definition, its findings will always be open to question. This chapter discusses the nature of this hypothetical-ness and reviews remedies to minimize its effects.[3]

In Section 4, I present a new method for overcoming problems of hypothetical-ness. This section reports the results of a real-money experiment used as part of a hypothetical valuation survey.

The survey used a willingness-to-accept (WTA) framework, which is the correct measure for many environmental problems but is often not undertaken because of poor results. In contrast, the WTA results reported here demonstrate "typical" valuation behavior. I argue that the real-money experiment helped the survey participants see valuation as a concrete exercise in which money and goods change hands. This experience in turn helped them understand what they were being asked to envisage in the more unfamiliar environmental valuation question.

Before tackling these issues, I first discuss valuation experiments in general. I describe how lessons from experimental economics have been used in environmental valuation. Although environmental valuation borrows freely from experimental economics, the exact connection between environmental valuation and valuation experiments has not been much explored. This connection is laid out in Section 2.

2. VALUATION EXPERIMENTS: DEFINITION AND A SHORT HISTORY[4]

Experimental economics and environmental valuation have grown up together, but their exact relationship has not been laid out. This section presents a short discussion of the use of experiments to value things.[5]

2.1 VALUATION EXPERIMENTS

Valuation experiments are a relatively new category of economic experiments, albeit with a long pedigree. A valuation experiment is one that elicits values as a tool for studying economic behavior. An example is the use of a Becker-DeGroot-Marshack mechanism (BDM) to elicit an individual's willingness-to-pay for a lottery (Becker, DeGroot, and Marshak 1964). Another example is the use of an auction to elicit individuals' compensation-demanded to taste the bitter substance SOA (Coursey, Hovis, and Schulze 1987).

What makes these two examples "valuation experiments" rather than some other kind of experiment is that the specific activity subjects engage in is the expression of a dollar value. What makes them "experiments" rather than simply "valuation exercises" is that their design allows researchers to study economic behavior, not simply record the value of the item being investigated. In a sense, valuation experiments attempt to understand behavior and values jointly, rather than separately.

Further examples of valuation experiments include the difference between willingness-to-pay and willingness-to-accept (Knetsch and Sinden 1984, Horowitz and McConnell 2002); the value of risk (Harless and Camerer 1994, Hey and Orme 1994); value of timing (Thaler 1981, Horowitz 1991, Harrison, Lau, and Williams 2002); endowment or reference point effects (Samuelson and Zeckhauser 1988, Horowitz, List, and McConnell 2005, Kahneman, Knetsch, and Thaler 1990); concavity of values (Horowitz, McConnell, and Quiggin 1999); and some public-goods provision experiments (Bohm 1972).[6] Hypothetical-versus-real experiments play a special role in the intersection of valuation experiments and environmental valuation (Cummings,

Harrison, and Rutström 1995, Cummings et al. 1997). Note that in all of these examples, the actual value of the item being studied is rarely the primary object of interest. On the other hand, the basic format of these experiments is the elicitation of a value.

Valuation experiments are a branch of what Davis and Holt (1993) identify as individual-choice experiments, which they define to include experiments on optimality, rationality, choice, and related topics. Among this group, the line between valuation experiments and non-valuation "choice experiments" will always be a shaky one because value and choice are so closely intertwined in economics.

I define valuation experiments as experiments that either (i) explicitly elicit a subject's willingness-to-pay or compensation-demanded for some well-defined item; or (ii) entail a choice that bounds the subject's value for some item. Thus, dichotomous choice, iterated dichotomous choice, polychotomous choice (choosing the best from a group of options), and ranking exercises constitute forms of valuation experiment. Although the subjects' values are mostly denominated in money terms, there are some cases in which values may be defined in non-money terms.

Early economic experiments were conducted using what are now known as "induced values." Subjects either received money or a token for their actions in the experiment; those actions had no value in and of themselves. As Harrison, Harstad, and Rutström (2002) note, early experiments were conducted almost entirely with induced values, although there are some important exceptions involving indifference curves, public goods, and choice under uncertainty (see Roth 1995.)

The alternative to induced values are "homegrown values." Homegrown means the subject comes to the experiment with the values rather than being assigned them as

part of the experiment; in other words, the values come from outside the experiment. The first application we can find of this term is Rutström (1998). Homegrown values are essentially identical to what environmental economists are interested in eliciting.

It should be obvious that most valuation experiments elicit homegrown values. It would be a mistake, however, to think that homegrown values are the only focus on valuation experiments, since several important valuation results have been derived from induced-value experiments. Irwin et al. (1998), for example, studied the BDM using induced values in order to test whether the BDM is truly incentive compatible. Other relevant induced-value experiments include Kagel, Harstad, and Levin (1987), Kagel and Levin (1993) and Noussair, and Robin and Ruffieux (2002).

2.2 History of valuation experiments

Valuation experiments are not a new enterprise. The famous Bernoulli paradox of 1738 is an early example of this kind of experiment: David Bernoulli asked his brother Nicholas his willingness-to-pay for a lottery that paid $n (in contemporaneous currency, of course) with probability $1/2^n$, where n ranged from 1 to infinity. The expected value of this lottery is infinity, but Nicholas was willing to pay only a finite amount, thus providing early evidence of risk aversion.[7]

This experiment already shows the amorphous mixture of valuation and behavior. Nicholas Bernoulli is being asked the value he places on the lottery; at the same time, he is demonstrating a behavior toward risk, by which I mean his predicted pattern of choices in other situations involving uncertainty. But it is clear that this exercise represents a valuation experiment rather than valuation itself. The experimenter is not truly interested

in the value of the lottery but in what that value reveals about attitudes toward risk. At the same time, the essence of the experiment is clearly the question about value, in this case, willingness-to-pay.

Other early examples came roughly 250 years later. In the early 1960's, Vickrey (1961) and then Becker, DeGroot, and Marshak (1964) examined methods to induce subjects to reveal their values. These were first applied to valuing lotteries rather than goods. It was several years later that experimental techniques were used for commodities or amenities. The earliest example I can find of a valuation experiment for an environmental commodity is the goose-hunting experiment in Bishop and Heberlein (1979).

2.3 History of Stated Preference Surveys

The previous section looked at how experimental methods came to be applied to environmental valuation. One might also ask how environmental valuation came to see the usefulness of the experimental literature. Stated preference surveys first developed independently of the experimental literature. Early stated preference surveys such as Hammack and Brown (1974) were conducted without any recourse to or acknowledgement of the experimental literature. Because valuation started out clearly as hypothetical, at about the time when real money became a crucial ingredient of experimental economics, the connection between experiments and environmental valuation was not immediately made. See Hanemann (1992) for discussion of the development of the stated preference approach to environmental valuation.

3. THE ROLE OF HYPOTHETICAL BIAS

Hypothetical bias is one of the most important contributions of valuation experiments to environmental valuation. Hypothetical bias means the difference between responses to hypothetical valuation experiments and real choices. Environmental valuation questions are typically considered hypothetical because the choice situations they pose are fabricated and because subjects' responses are not directly tied to the environmental decision ultimately made. The ultimate purpose of environmental valuation, however, is to recover true values. Thus, understanding potential discrepancies between hypothetical and real responses is at the very center of research for valuation experiments.[8]

Before tackling this issue, however, it is worth asking to what degree stated preference surveys are indeed hypothetical. This is the goal of Section 3.1.

3.1 ARE STATED PREFERENCE SURVEYS HYPOTHETICAL?

Most analysts discuss contingent valuation surveys as if they were hypothetical. Surveys are hypothetical presumably because the vast majority of environmental decisions are not made on the basis of a benefit-cost analysis (see, for example, Morgenstern 1997). A large proportion of benefit-cost analyses are conducted for regulations that specifically disallow cost considerations or, more generally, that dictate specific non-benefit-cost decision criteria; many benefit-cost analyses are conducted for regulations that have already been issued. In such situations, the contingent valuation survey must be considered hypothetical because the specified environmental decision is unconnected to subjects' responses.

In most cases, hypothetical-ness is obvious from the wording and the situation described, as in this question from Jones-Lee, Hammerton, and Phillips (1985):[9]

> Imagine that you have to make a long coach trip in a foreign country. You have been given £200 for your traveling expenses, and given the name of a coach service which will take you for exactly £200. The risk of being killed on the journey with this coach firm is 8 in 100,000.
>
> You can choose to travel with a safer coach service if you want to, but the fare will be higher, and you will have to pay the extra cost yourself.
>
> (a) How much extra, if anything, would you be prepared to pay to use a coach service with a risk of being killed of 4 in 100,000 – that is half the risk of the one at £200?

Furthermore, even in cases that are much more realistic, this hypothetical-ness remains an obvious feature of the valuation scenario:

> Because everyone would bear part of the cost, we are using this survey to ask people how they would vote if they had the chance to vote on the program…
>
> At present, government officials estimate the program will cost your household a total of $50. You would pay this in a special one time charge in addition to your regular federal taxes. This money would only be used for the pgram to prevent damage from another large oil spill in Prince William Sound.
>
> If the program cost your household a total of $50 would you vote for the program or against it?
>
> (Main interview questionnaire, Alaska Exxon Valdez Study (Westat, undated))

Carson, Groves, and Machina (2002, hereafter CGM) argue that many of these situations should be considered non-hypothetical despite the wording and despite the absence of a concrete connection to environmental policy-making. Their argument runs as follows: Suppose a subject is asked whether her household would be willing to pay $50 for a conservation policy that would preserve spotted owl habitat. Suppose there were no concrete connection between the subject's response and the U.S. government's spotted owl policies. CGM argue that despite the vagueness of the link between subjects' responses and the policy decision, it is reasonable to assume that there is at least some

probability that the subject's response will tilt the policy in the direction described. That is, if the subject says yes, then it is more likely that a policy costing $50 and improving spotted owl habitat will be enacted. If she says no then the policy is less likely to be enacted. Thus, the subject has the incentive to answer truthfully. See also Cummings et al. 1997.

There are two important caveats. First, the subject must believe that the survey accurately describes the costs and environmental quality choices, as CGM note. This condition is almost surely violated by even the best contingent valuation survey. For example, suppose the policy turns out to cost $100 (and this is known before the final regulatory decision is made.) If this subject's response (to the cost of $50) is used to make a decision about the $100 policy, then the "true cost" incentive compatibility condition is violated. Likewise, if the subject's $50 response will be used to make decisions about policies other than the one being described, then the "true environmental quality" incentive compatibility condition is violated.

All of the common estimation techniques for closed-ended questions introduce the first kind of violation. The reason is that they estimate a distribution of responses rather than using the sample responses at each cost level when predicting population responses at that cost level. That is, a subject's response at a cost of $50 is, through the estimation procedure, part of the econometrician's inference about her responses at all other cost levels. This violates the condition that the reported cost is the true cost, because a subject's discrete-choice response (under a true cost scenario) would provide information only about her willingness to pay that cost. Open-ended responses incur similar objections; see Section 4.2.

Benefits-transfer introduces the second kind of violation. Under benefits-transfer, a subject's response about the spotted owl conservation policy might be used to infer her willingness-to-pay for other endangered species protection policies or even her willingness-to-pay for environmental actions in general. Any respondent who believes that her response may be used for these kinds of inferences will adjust her response accordingly. Note that respondents would be perfectly justified in such a belief.[10]

The second caveat is that even when the scenario is accurate and therefore incentive-compatible, it surely matters, from a practical standpoint, how much more likely it is that the policy will be enacted or not based on the subject's response.

If the actual policy choice will be determined by subjects' responses, as in a binding referendum, then a respondent has a very strong incentive to answer truthfully. If the actual decision rests on many pieces of information (as is generally the case), and the subject's response makes it "just a little more likely" that the policy will be enacted or not, then she has a much weaker incentive to answer truthfully. While both of these cases will lead individuals to reveal their true willingness-to-pay according to the theoretical model of CGM, these cases may not lead to the same responses in practice.

Whether these are big violations remains an empirical matter. Valuation experiments would seem to be particularly useful in this regard. (For some evidence, see Taylor et al. 2001).

3.2 EXPERIMENTAL APPROACHES TO HYPOTHETICAL BIAS

Murphy and Stevens (2004) identify three approaches to hypothetical bias in the literature:

(1) Calibration. Researchers can measure the ratio of hypothetical to real values in experiments that meet the test of revealing true willingness to pay. Calibration means dividing values from stated preference surveys by some number – 2, for example – in order to arrive at the true value. Calibration can attempt to correct for: (a) respondent demographic characteristics, such as income or whether the respondent is a student; (b) respondent certainty about his or her stated value; (c) survey features; or (d) valuation question. For recent studies of calibration, see Hofler and List (2004) and List and Shogren (2002).

Murphy and Stevens (2004) note that some studies have implied that calibration could be tailored to payment amounts. For example, consider a closed-ended survey. The proportion of subjects who say yes to a real and hypothetical question at the $5 level may be roughly the same (calibration = 1); but, the proportion who say yes at the $50 level may be higher for hypothetical than for real questions (calibration > 1). Thus, with sufficient experimental information, calibration could be tied to the payment amount.

This experimental finding is unusable for calibration, however, since it is impossible to know ex ante whether a given payment level is low or high. Put another way, the decision to label a payment amount low or high, and therefore subject to a specific calibration factor, is necessarily based on other information or assumptions such as the income effect or the "reasonableness" of the payment amount. This information should either be used explicitly, in the case of the income effect, or not at all.

By extension, calibration also cannot be conditioned on a subject's reported value in an open-ended survey. For example, it may be tempting to assume that subjects who say they are willing to pay $5 in a stated-preference survey are probably telling the truth

whereas those who say they are willing to pay $500 are probably not. Indeed, this assumption will often be correct. But no calibration can be formulated from this relationship. Calibration on the payment amount would be due either to the respondent's income, in which case calibration should be treated as a demographic calibration, as in approach (a) above. Or, calibration would be due to the researcher's ex ante belief about the ballpark value of the item being assessed. This belief cannot legitimately be used when doing valuation.

(2) Survey design features. Researchers can use specific survey design features to minimize hypothetical bias.

When analysis identifies survey features that have calibration levels close to one, an alternative to calibration is to directly incorporate those features in a stated-preference survey. In this case, the two approaches to hypothetical bias, calibration and survey design, are essentially identical. In general, a good survey design feature is one with a calibration ratio close to one.

Survey design has a broader agenda, however. It refers more broadly to the use of design features that are intellectually appealing but that cannot be reliably calibrated; in other words, constructive validity. Experimental evidence may be either impossible to provide, or currently incomplete or qualitative rather than quantitative. I follow this topic further in Section 3.3.

(3) Valuation questions. Researchers can reframe the "valuation exercise" itself. An example is the use of closed-ended rather than open-ended questions. This is fundamentally different from survey design, which refers to how a specific valuation exercise, such as an open-ended survey, is presented. In the case of open- vs. closed-

ended questions, the former elicits an individual value; the latter only bounds that value and provides, with additional assumptions, a population average value.

3.3 CONTROL OF AN EXPERIMENT'S "REAL-NESS"

The benefit of experiments is the control that they afford. Their particular benefit for illuminating hypothetical bias is that they allow the researcher to control the nature and degree of hypothetical-ness in a valuation question. In a non-experimental setting, it is difficult to know how real a valuation question is perceived to be. But in the experimental context it is possible to construct the valuation exercises explicitly as hypothetical or real. This is a necessary step because a quantitative understanding of the effects of survey design features requires a clear measure of the degree to which a valuation task is perceived to be "real."

Unfortunately, there are serious limitations on the ability of experiments to provide relevant controls on the real-ness or hypothetical-ness of a valuation experiment.

Harrison, Harstad, and Rutström (2002) point out that hypothetical bias is almost always studied for private goods. They note that even in a laboratory experiment the outside market for these goods (that is, the ability to purchase or sell an item outside the experiment) will likely play an important role but beliefs about this outside market are unobservable. Note that because these are private goods, those outside markets necessarily exist. Several experimental papers have attempted to control outside market beliefs, with varying success (e.g., Shogren et al. 1994). Horowitz and McConnell (2000) measured the effect of dropping experimental willingness-to-accept observations

that deviate too greatly from resale values of items. They cannot, however, determine what the "right" role for resale is.

The problems for environmental goods are even greater. First, very few research projects have successfully created a real valuation experiment on the scale that is needed for reliable inference. A "real" experiment involving environmental policy is impossible almost by definition, since those policies must be chosen based on legal procedures that preclude binding valuation exercises. The outstanding exception to this claim is the California referendum study of Carson, Hanemann, and Mitchell (1986).

Experiments in which subjects provide funds for a public good that will actually be purchased (Brookshire and Coursey 1987, Cummings et al. 1997, Landry et al. forthcoming) overcome some of these objections, but other problems remain. The unobservable-outside-options problem remains, since in most cases subjects could donate outside the experiment to the public good. The number of participants is considerably smaller than would be affected by most environmental policies. Many experiments (although not Landry et al. forthcoming) provide subjects with a participation payment, which likely contaminates the experiment. Perhaps most importantly, fund-raising for public goods is a very different kind of public choice from that posed by most environmental policies.

Many dimensions that are important for real-ness have not yet been the focus of survey design, and it will be difficult to make them so. As Section 3.1 makes clear, it is important that both the cost and environmental quality change being described be concrete and immutable. Yet very few valuation studies focus on these dimensions.

In summary, one of the strongest potential contributions of valuation experiments is to elucidate hypothetical bias and thus improve the reliability of hypothetical environmental valuation questions. However, this task requires that experimenters control the "real-ness" of their treatments. This is intrinsically difficult for environmental valuation because (i) realistic real environmental valuation situations are extremely difficult to construct; and (ii) the nature of environmental regulation often precludes a clear and decisive role for valuation, which is essential for real valuation. Environmental valuation surveys are "not real" in ways that are difficult for experiments to treat or overcome.

3.4 THE "BUDGET CONSTRAINT" PROBLEM

The previous sections have argued that experiments are integral to improving valuation surveys. However, experimental methods to reduce hypothetical bias in environmental surveys are more difficult to devise than has generally been recognized.

The most promising alternative approach is based on construct validity. By construct validity, I mean the adoption of techniques that are intellectually and intuitively valid. Direct empirical evidence is impossible, but indirect empirical is valuable and essential.

Construct validity is most useful for the problem of designing surveys to encourage subjects to take their budget constraints seriously. The goal is to have survey respondents use the same mindset that they use in making familiar spending and savings decisions involving their own money. This mindset is unobservable, of course, and therefore it is difficult, if not impossible, to test whether subjects are "taking their budget constraints seriously."

Obvious departures from budget-mindfulness can sometimes be spotted; for example, Horowitz and McConnell (2000) observed some $1 million valuations of binoculars that cost $25. Such a response is clearly wrong. But there is no obvious line to draw between reasonable answers and unreasonable ones, and any attempt to impose a standard of reasonableness should be resisted.

On the other hand, sharp researchers can recognize whether the survey design (as opposed to its results) adequately encourages participants to take their task seriously. This is the role for construct validity. Survey design is as much art as science; this is the art part.

From the inception of contingent valuation, economists have used a wide range of reminders to help put subjects in the appropriate frame of mind. Tolley and Babcock (1986) used the following approach when valuing the benefits of clean air:

> Before we start, please look at this card showing how a typical family spends its take-home income.
>
> When you pay to avoid symptoms, the money will have to come out of one of the categories shown. We'll leave the card here so that you can think about where the money comes from that you would spend to avoid the symptoms.

The experiment shown in Section 4 is proposed with this approach in mind. Subjects who make a real-money decision that takes the same form as the valuation decision are more likely to adopt the approach that economists hope they would.

4. THE USE OF A REAL-MONEY EXPERIMENT IN A STATED-PREFERENCE SURVEY

This section proposes a technique for improving valuation surveys. A real-money experiment is conducted as part of the survey, before the key valuation question. Its

purpose is to use the real-money, real-goods experience to help put survey subjects in the right "frame of mind" for the environmental valuation question.

This procedure requires that the valuation survey take place in-person. Neither mail nor telephone surveys are suitable for this procedure. The procedure works best when the valuation survey is administered to a group of subjects simultaneously. The group format approach is discussed in Section 4.2.

4.1 PROCEDURE

The valuation experiment takes the form of a willingness-to-accept public-choice experiment. The experiment uses an open-ended format and a median-value rule.

Each of the subjects, seated in a room together, is given a small item such as a mug, flashlight, or pair of binoculars. Subjects are asked, individually, the minimum payment they would require to return this item to the experimenter. A public-choice variant of the Becker-DeGroot-Marschak mechanism is used to determine whether subjects will keep their items or return them to the experimenter and, if returned, the amount of the payment.

The experimenter draws a random per-item price, called the offered amount. If more than 50 percent of the subjects are willing to surrender their item for that amount, then all subjects must return their items to the experimenter and all subjects are paid the offered amount. The payment is made on the spot, in real money.[11] If less than 50 percent of the subjects are willing to surrender their item for that price, then all subjects keep their items and no payment is made. A sample sheet is shown in Figure 1.

This is essentially a standard willingness-to-accept experiment, but here used for a collective choice. The experiment highlights for participants both (i) the valuation exercise (each participant must ask himself or herself, "What is this item really worth to me?") and (ii) the public choice decision, namely, how individuals' values are used to make a collective choice. The purpose of these components is to help participants get in the mindset of making value-based decisions about public goods.

4.2 DISCUSSION

4.2.1 Group presentation format

The group presentation format is useful for making clear the public-choice dimension of environmental valuation. In the absence of a group format, experiments could use a standard BDM.

The group format has long been the norm in behavioral experiments. But its use specifically for valuation (that is, for experiments that might otherwise by conducted one-on-one) is still not established. The group format was first suggested for contingent valuation (to my knowledge) by Richard Carson. Following Carson's recommendation, Horowitz, McConnell, and Quiggin (1999) used this format to conduct private goods valuation at lower cost to the experimenter than one-on-one valuation. John List has conducted many field experiments using the group format to study both private goods and public goods. Other examples are Cummings et al. (1997). Schelling has long advocated using group discussion as part of the (individual) valuation exercise. His reasoning is that most real-world opinions are formed and expressed in open discussion.

4.2.2 Willingness-to-accept

This experiment requires a willingness-to-accept format, which is much less common than willingness-to-pay. The alternative willingness-to-pay format is essentially impossible because experimenters cannot compel subjects to pay for a public good. One remedy, an up-front payment to subjects (e.g., $10, with the specified payments being less than $10), is highly problematic. If the experimenter can compel all subjects to pay for a public good, then that money cannot also be considered the subjects' to spend as they please, which is the key feature of a real money valuation experiment!

The willingness-to-accept format is not a problem for the experiment but for the subsequent environmental valuation survey. If the valuation survey is framed in terms of willingness-to-pay, some of the lessons of the real-money experiment will be lost, or willingness-to-accept. If the valuation survey is framed in terms of willingness-to-accept, then the experiment's lessons are not lost. However, environmental valuation using willingness-to-accept will almost surely lead to non-credible value estimates.

4.2.3 Mean vs. median willingness-to-pay

This experiment also requires a median-value approach. Note that many of the real-money public goods experiments such as Cummings et al. (1997) or Horowitz, List, and McConnell (2005) also adopt a median-value or median-voter approach. For an exception, see Brookshire and Coursey (1987). The median value approach is incentive compatible under a broad range of conditions. In contrast, mechanisms using total or average willingness-to-pay are not incentive compatible even if they are derived from closed-ended surveys. See CGM for a review this literature as it applies to valuation.

Many economists might hope that the total-value approach could be used despite its not being truly incentive compatible. There is a lesson, however, in this body of experimental work using the median value. In a real-world experiment (such as Cummings et al. 1997) it is necessary to explain to subjects how their responses will be used to make the public goods decision. For a closed-ended survey in which subjects are assigned different cost amounts, the required explanation is complicated and therefore probably not feasible. Such a mechanism would be difficult to administer even if subjects did not take into account its non-incentive-compatible opportunities.

On the other hand, for an open-ended survey the required explanation would be simple. But subjects could easily understand how to exploit the non-incentive-compatible loopholes, therefore making the open-ended total-value approach also unworkable.

This paper's proposed experiment, because it is open-ended and incentive compatible, allows economists to observe individual values. This is advantageous for experimental purposes. But this information cannot be used for actual valuation decisions unless the experimenter is willing to be untruthful to the subjects.

It is possible to conduct this experiment with a closed-ended (yes-no) survey. The question format would be the standard dichotomous choice format. The decision rule would be that after the offered amount is drawn, the experiment looks to see whether 50 percent of the responses at that amount would vote in favor of accepting the offer.

The open-ended survey shown in Figure 1 is highly desirable from an experimental point of view. A closed-ended survey might be valuable, however, if the subsequent valuation survey was closed-ended.

4.2.4 Experiments used in valuation surveys

The literature on stated preference surveys that used real experiments on unrelated items as part of the survey is relatively slim.

4.3 RESULTS

I conducted, with K.E. McConnell, a real-money individual-choice experiment followed by a hypothetical valuation survey. The real-money experiment is described in Horowitz, McConnell, and Quiggin (1999). It is similar to the public-choice experiment shown in Figure 1, but the rule about whether the individual kept his item or returned it for money was made at the individual level, not the group level. We did not, however, conduct a public-choice experiment, nor did we conduct a separate treatment that included a valuation survey without the preceding real-money experiment. Thus, the results described here do not fully conform to the research design described in Section 4.1.

The hypothetical valuation survey examined values for wetlands preservation using a willingness-to-accept framework. The survey instrument is shown in Figure 2. It used a one-shot closed-ended format with four possible values for the payment amount: $48, $100, $150, and $300.

Results are shown in Table 1. The subjects were members of a local Parent-Teacher Association. Fifty subjects took part.

4.3.1 Discussion

I used the data in Table 1 to estimate a mean willingness-to-accept of $88. There were roughly 1.9 million Maryland households in 1996, which yields a total WTA of $167.2 million. This works out to $4644 per acre.

Note that willingness-to-accept experiments are often poorly behaved, so this pattern of responses represents a small victory of experimental technique.

Based on results in Horowitz and McConnell (2002) for public goods, the predicted mean willingness-to-pay would be $8.45 (= $88/10.4).

There is one word of caution, however. These results came after several valuation exercises to assess willingness-to-accept reduced protection of sea turtles (using a different subject pool.) Protection of sea turtles turned out to be such a high-value issue that the great majority of respondents reported that there was no compensation that they would accept for a diminution of sea turtle populations (even though the survey stated that turtles would not become endangered as a result of the action.) Analysis was essentially impossible because there were so few "yes" responses.

There are two interpretations of the turtle results in comparison with the "well-behaved" wetlands results. First, it is possible that turtle values are indeed extremely high, in which case the valuation survey should be considered a success. But it is also possible that turtle results are not stratospherically high and that the accompanying real-money experiment was not sufficient to overcome the problem of unfamiliarity that often makes environmental valuation unworkable.

5. CONCLUDING REMARKS

This chapter has two themes. First, valuation experiments are a separate category of experiment, to be distinguished from other experiments. Valuation experiments elicit preferences, and from these experiments we learn both about preferences and how individuals express those preferences. This is the very essence of economics.

This joint inquiry into preferences and the expression of those preferences is one of the paramount contributions that environmental economics, together with experimental economics, has made to the economics profession. Environmental economists were among the first economists to recognize the importance of the question, "What do people really care about and how do we uncover this?" and to take it seriously as an empirical matter.

Second, this chapter has shown how a real-money experiment can be used as part of a stated-preference survey. The real-money experiment provides a useful "decision context" for subjects without polluting the stated-preference survey. This chapter did not attempt to test hypotheses about this combination of experiments. I leave this for future research.

Table 1. Wetlands values – Raw data

	Payment offer			
	$48	**$100**	**$150**	**$300**
Yes	2	2	1	6
No	12	12	8	7
(Total subjects)	14	14	9	13
Percent Yes	14%	14%	11%	46%

Figure 1. Real-money public-choice valuation experiment

Flashlight Values

Each person in the room has been given a flashlight. You will either get to keep your flashlights or return them to the administrator, in which case everyone will receive some payment. The amount of the possible payment will be determined at the end of the exercise.

What is the smallest payment that would make you on-the-fence between keeping your flashlight and returning it to the administrator?

Amount: _____

This is like a vote to keep or return your flashlight. Suppose I offered you 10 cents more than the amount you write down. Would you definitely vote ***for*** accepting the offer, if you knew that was the amount I was offering?

Suppose I offered 10 cents less than the amount you write down. Would definitely vote ***against*** accepting the offer?

At the amount you write down, you should be ***on-the-fence*** about voting for or against accepting the offer.

There is no right or wrong answer; it just depends on how much you like the flashlight.

How the Outcome will be Determined:

1. After everyone has turned in his sheet, I will put all the sheets in order, with the amounts going from lowest to highest.

2. I will draw the payment amount randomly, using the device at the front of the room.

3. Once we know the amount, I will determine whether at least half of the people would have voted for or against accepting that amount.

4. If the amount I am offering is high enough (more than half the people would have voted *for* it), then everyone will return his or her flashlight and receive the offered amount.

5. If the amount I am offering is too low (more than half the people would have voted *against* it), then everyone will keep his or her flashlight and receive no money.

Figure 2. Environmental valuation survey

H. Wetlands Acreage

One environmental issue that society faces is the preservation of *wetlands*. Wetlands are marshes, forested swamps, and other areas that are under water part of the year. Wetlands vary from less than an acre in size to thousands of acres. Wetlands provide many services to the natural environment such as maintaining water quality and providing habitat for fish, birds, frogs, raccoons, and other animals.

Unfortunately, there is competition for the use of wetlands. Wetlands can be drained and used for farms, for houses, or for industries that create jobs and raise incomes.

To preserve wetlands and protect the environment, the state of Maryland purchases wetlands and places them in parks and preserves. There are approximately 300,000 preserved wetlands acres on the eastern shore.

State budget analysts have made a suggestion about some of these wetlands. They have suggested selling 36,000 acres to the highest bidder. If this land is sold, whoever buys it will be free to do with it as he wishes, subject to other state laws. The area will probably be developed.

If these wetlands acres are not sold, there will be 300,000 preserved wetlands acres on the eastern shore. If the 36,000 acres are sold, there will be 264,000 preserved wetlands acres, but there will also be more money available to the state government. This money can be used for schools or police or used to cut state income taxes.

The 36,000 acres are the same kind of wetland as the other 264,000 acres. They do not have any distinguishing environmental characteristics. The acres were chosen to be considered for would be a one-time payment that every household in the state would receive if the wetlands were sold.

H. Wetlands Acreage, cont.

Name: _____

The purpose of this survey is to determine which outcome you prefer.

Suppose that we that we were offering to write you a check for $150. This represents how much money the state would receive, per household, by not keeping these acres. We would like to know whether you prefer: (i) a check for $150, and keeping 264,000 acres; or (ii) no check, and keeping 300,000 acres.

The decision for you is similar to the flashlight survey. The main difference is that now we can tell you ahead of time the dollar amount that is being offered. This would be a one-time payment that every household in the state would receive if the wetlands were sold.

The outcome will depend on whichever option receives the most votes.

QUESTION H:

Please indicate which outcome you would prefer:

_____ A check for $150, and keeping 264,000 wetland acres; or

_____ No check, and keeping 300,000 wetland acres.

The details and the situation in this question are hypothetical.

Our purpose is to explore a "what if" scenario.

6. REFERENCES

Becker, G., M. DeGroot and J. Marschak (1964), 'Measuring Utility by a Single-Response Sequential Method,' Behavioral Science, 9, 226-36.

Bishop, R.C. and T. Heberlein (1979), 'Measuring Values of Extramarket Goods: Are Indirect Measures Biased?' American J. Agricultural Economics, 61, 926-30.

Bohm, P. (1972), 'Estimating demand for public goods: An experiment,' European Economic Review, 3, 111-30.

Brookshire, D.S. and D.L. Coursey (1987), 'Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures,' American Economic Review, 77, 554-66.

Carson, R., W.M. Hanemann and R.C. Mitchell (1986), 'Determining the demand for public goods by simulating referendums at different tax prices,' Unpublished manuscript.

Carson, R., T. Groves and M. Machina (2002), 'Incentive and Informational Properties of Preference Questions,' Unpublished manuscript.

Coursey, D.L., J.L. Hovis and W.D. Schulze (1987), 'The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value,' Quarterly Journal of Economics, 102, 679-90.

Cummings, R., S. Elliott, G.W. Harrison and J. Murphy (1997), 'Are Hypothetical Referenda Incentive Compatible?' Journal of Political Economy, 105, 609-21.

Cummings, R., G.W. Harrison and E. Rutström (1995), 'Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible,' American Economic Review, 85, 260-66.

Davis, D.D. and C.A. Holt (1993), Experimental Economics, Princeton University Press.

Epstein, R.A. (2003), 'The regrettable necessity of contingent valuation,' *Journal* of Cultural Values, 27, 259-274.

Hammack, J. and G.M. Brown (1974), Waterfowl and Wetlands: Toward Bioeconomic Analysis, Johns Hopkins University Press for Resources For the Future.

Hanemann, W.M. (1992), 'Preface,' in Pricing the European Environment, S. Navrud (ed.), Scandinavian University Press.

Hanemann, W.M. (1994), 'Valuing the Environment through Contingent Valuation,' Journal of Economic Perspectives, 8, 19-43.

Harless, D. and C. Camerer (1994), 'The Predictive Utility of Generalized Expected Utility Theories,' Econometrica, 62, 1251-89.

Harrison, G.W. (1992), 'Theory and Misbehavior of First-price Auctions: Reply,' American Economic Review, 82, 426-43.

Harrison, G.W., R. Harstad and E.E. Rutström (2002), 'Experimental Methods and Elicitation of Values,' Unpublished manuscript.

Harrison, G.W., M. Lau and M. Williams (2002), 'Estimating Individual Discount Rates for Denmark: A Field Experiment,' American Economic Review, 92, 1606-17.

Harrison, G.W. and J.A. List (2004), 'Field experiments,' Journal of Economic Literature, 42, 1009-55.

Hey, J.D. and C. Orme (1994), 'Investigating Generalizations of Expected Utility Theory Using Experimental Data,' Econometrica, 62, 1291-1326.

Hofler, R. and J.A. List (2004), 'Valuation on the frontier: Calibrating actual and hypothetical statements of value,' American Journal of Agricultural Economics, 86, 213-21.

Horowitz, J. (1991), 'Discounting Money Payoffs: An Experimental Analysis' in S. Kaish and B. Gilad (eds.), Handbook of Behavioral Economics Vol. 2B, JAI Press.

Horowitz, J., J.A. List and K.E. McConnell (2005), 'Diminishing Marginal Value,' Unpublished manuscript.

Horowitz, J. and K.E. McConnell (2002), 'A Review of WTA/WTP Studies,' Journal of Environmental Economics and Management, 44, 426-47.

Horowitz, J. and K.E. McConnell (2000), 'Values Elicited from Open-Ended Real Experiments,' Journal of Economic Behavior and Organization, 41, 221-37.

Horowitz, J., K.E. McConnell and J. Quiggin (1999), 'A Test of Competing Explanations of Compensation Demanded,' Economic Inquiry, 37, 637-46.

Irwin, J., G. McClelland, M. McKee, W. Schulze and N. Norden (1998), 'Payoff dominance vs. cognitive transparency in decision making,' Economic Inquiry, 36, 272-85.

Jones-Lee, M.W., M. Hammerton and P. Philips (1985), 'The Value of Safety: Results of a National Sample Survey,' Economic Journal, 95, 49-72.

Kagel, J. (1995), 'Auctions: A Survey of Experimental Research,' in J. Kagel and A. Roth (eds.), Handbook of Experimental Economics, Princeton University Press.

Kagel, J., R. Harstad and D. Levin (1987), 'Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study,' Econometrica, 55, 1275-1304.

Kagel, J. and D. Levin (1993), 'Independent Private Value Auctions: Bidder Behaviour in First-, Second-, and Third-Price Auctions with Varying Numbers of Bidders,' Economic Journal, 103, 868-79.

Kahneman, D., J. Knetsch and R. Thaler (1990), 'Experimental Tests of the Endowment Effect and the Coase Theorem,' Journal of Political Economy, 98, 1325-48.

Knetsch, J. and J. Sinden (1984), 'Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value,' Quarterly Journal of Economics, 99, 507-21.

Landry, C., A. Lange, J.A. List, M.K. Price and N. Rupp (forthcoming), 'Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment,' Quarterly Journal of Economics.

List, J.A., and C. Gallet (2001), 'What experimental protocol influence disparities between actual and hypothetical stated values?' Environmental and Resource Economics 20, 241-54.

List, J.A. and J. Shogren (2002), 'Calibration of Willingness-to-Accept,' Journal of Environmental Economics and Management, 43, 219-33.

Morgenstern, R. (1997), Economic Analyses at EPA, Resources for the Future.

Murphy, J.J., P. Allen, T.H. Stevens and D. Weatherhead (2005), 'A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation,' Environmental and Resource Economics, 30, 313-25.

Murphy, J.J. and T.H. Stevens (2004), 'Contingent Valuation, hypothetical bias, and experimental economics,' Unpublished manuscript.

Noussair, C., S. Robin and B. Ruffieux (2002), 'Revealing consumers' willingness-to-pay: A comparison of the BDM mechanism and the Vickrey auction,' Unpublished manuscript.

Roth, A. (1995), 'Introduction to Experimental Economics,' in J. Kagel and A. Roth (eds.), Handbook of Experimental Economics, Princeton University Press.

Rutstrőm, E.E. (1998), 'Home-grown Values and Incentive Compatible Auction Design,' International J. of Game Theory, 27, 427-41.

Samuleson, W. and R. Zeckhauser (1988), 'Status Quo Bias in Decision Making,' Journal of Risk and Uncertainty, 1, 7-59.

Shogren, J., S. Shin, D. Hayes, and J. Kliebenstein (1994), 'Resolving Differences in Willingness to Pay and Willingness to Accept,' American Economic Review, 84, 255-70.

Taylor, L.O., M. McKee, S. Laury, and R. Cummings (2001), 'Induced-value tests of the referendum voting mechanism,' Economics Letters, 71, 61-5.

Thaler, R. (1981), 'Some Empirical Evidence on Dynamic Inconsistency,' *Economics* Letters, 8, 201-7.

Tolley, G.S., and L. Babcock (1986), 'Valuation of Reductions in Human Health Symptoms and Risks,' U.S. Environmental Protection Agency, Office of Policy Analysis.

Vickrey, W. (1961), 'Counterspeculation, Auctions, and Competitive Sealed Tenders,' Journal of Finance, 16, 8-37.

Westat, Inc. (undated), National Opinion Survey (Alaska Exxon Valdez Study), http://econ.ucsd.edu/%7Ercarson/AKsurvey.pdf.

Footnotes

[1]I thank John List and Ted McConnell for helpful comments.

[2]This chapter does not address the issue of why environmental valuation should be considered "essential"; for this, see Hanemann (1994) or Epstein (2003), for example. I also leave out the valuation of ecological services, which is better thought of as a form of market, rather than nonmarket, valuation.

[3]The literature on hypothetical bias is vast; see List and Gallet (2001) or Murphy et al. (2005) for reviews.

[4]In writing this chapter I benefited greatly from the well-written and amazingly comprehensive Handbook of Experimental Economics, especially Chapter 1 by Al Roth (1995) and Chapter 7 by Kagel (1995).

[5]Harrison and List (2004) discuss the definition of an experiment.

[6]A full list of valuation experiments is beyond the scope of this chapter. In each case, I have tried to give a seminal article and/or review.

[7]One of the curious things about this famous experiment is that the actual value response is not typically reported. Whether the respondent gave an actual money value or merely a qualitative response is not typically reported.

[8]The term hypothetical bias is a bit misleading, but like "contingent valuation," it has stuck. If analysts are able to calibrate actual willingness-to-pay based on hypothetical survey responses, then the goal of survey design should be precision, not accuracy. Of course, accurate calibration is difficult, so survey design must focus on both precision and accuracy.

[9]An amusing feature of this question is its framing of the safety decision as taking place "in a foreign country."

[10]There is a useful opinion-poll analogy: Suppose a law is up for consideration and a member of the public is polled about whether he supports the law. The respondent has an incentive to respond truthfully even though he has no say in whether the law is passed because no matter how small his voice, his opinion must sway the vote in the right direction. On the other hand, suppose that a general issue is being debated and there are many potential remedies being considered, and that an opinion poll is conducted about a single concrete proposal to address the issue. Then the respondent will answer the question based on how he

thinks his response will shape the general debate or on what he thinks the "real" issue is, not simply the specific proposal being asked about.

[11]In past experiments, I have simply written checks to all subjects in the event that their responses led them to be paid the offered amount. At the end of the experiment, all subjects will have either a flashlight (or similar item) or a check for a small amount of money. Harrison (1992) argues that the BDM is a weak instrument for eliciting values.