Some Issues in Continuous- and Discrete-Response Contingent Valuation Studies

W. Michael Hanemann

I want to begin with a caveat. I have no training in psychology, and my interest in the field is that of an amateur. Thus, the confidence with which I shall make statements about psychological facts is in inverse proportion to my expertise. My point of departure is the assumption that the most serious problem with contingent valuation (CV) methodology is hypothetical bias. Apart from introspection, the evidence for this assumption comes from two classic studies conducted by Richard Bishop and his colleagues at the University of Wisconsin. The first study by Bishop and Heberlein (1979) involved a survey of two random samples of individuals who had obtained permits to hunt ducks in the Horizon Zone of East Central Wisconsin. One sample group was asked the hypothetical question whether they would be willing to sell their permit for a specified sum of money (this sum differed among individuals). The other sample group was given a real opportunity to sell their permit for a specified sum of money; of 237 hunters surveyed, 105 actually sold their permits.

Using a methodology which I discuss in a later section, Bishop and Heberlein estimated willingness-to-sell (WTS) functions from these two sets of data. From the fitted WTS functions, they derived an estimate of the expected selling price for an average individual which I denote C'. In Hanemann (1984), which is summarized in the third section, I propose two other measures of the average individuals' selling price which I denote C* and C^+. The results are:

	C'	C*	C^+
Hypothetical sale	\$101	83.16	∞
Real sale	\$ 63	31.02	114.22

W. Michael Hanemann is an associate professor of agricultural and resource economics at the University of California, Berkeley.
* Giannini Foundation Paper No. 322 (reprint identification only).

As might be expected, in the hypothetical "market" people exaggerate the amount of money required to induce them to sell their duck hunting permit. The question is: Are the differences in WTS statistically significant? Since the values of C', C*, and C^+ are all derived from the estimated WTS functions, an appropriate procedure is to test whether the coefficients of the two WTS functions are significantly different—this is equivalent to computing confidence intervals for the two estimates of C', C*, and C^+ and testing whether they overlap. The likelihood ratio statistic for testing the null hypothesis that the two WTS functions are the same is 25.14 while the .05 critical value of the relevant $X^2(2)$ distribution is 5.99. It seems clear, therefore, that the participants in the hypothetical market experiment responded in a significantly different manner from the participants in the real market experiment.

The second study by Boyle, Bishop, and Welsh (forthcoming) deals with the problems of starting point bias in conventional, iterative bidding CV experiments. Two samples of individuals were surveyed and asked about their willingness to pay to obtain a permit for a one-day deer hunt at Sandhill Wildlife Demonstration Area in Wisconsin. In one case this was a hypothetical question, and the participants did not actually have the opportunity to buy a permit; in the other case this was a *real* question, and the participants could actually buy a permit. In each case an iterative bidding format was employed, but the starting points varied widely among the participants in each survey. The resulting final "bids" were regressed on the starting point values. In the case of the hypothetical "market," there was a statistically significant positive correlation between the final bid and the starting value. This suggests that responses can be manipulated by the choice of the starting point: You can induce higher valuations by using higher starting points. However, in the real market experiment, there was *no* significant correlation between the starting points and the final bids; when people were faced with a real opportunity to buy a hunting permit, their expressed willingness to pay was independent of the starting point. This, too, suggests that participants in hypothetical market experiments behave differently than participants in real market experiments.

At this point it may be useful to offer a definition of hypothetical bias, based on a homemade theory of psychology. I want to suggest that, most of the time, people do not consciously *know* their own preferences; they usually cannot introspect their utility functions. Instead, they discover their preferences when they actually make a choice: a decision "pops into" their head. Their preferences are revealed to them as part of the actual choice. However, preferences are fairly stable (there may be a random component, but there also is a substantial deterministic component); therefore, if a person has faced the same type of choice on several previous occasions, he can *estimate* his own preferences with reasonable accuracy—he can predict what he would do if the choice arose in the future—by observing his own past decisions. As in economic theory, we rely on revealed preference to infer our own utility function and predict our behavior. Consequently, if a person is asked how much he would be willing to pay for an item which he has, in fact, considered buying in the past, he can give a fairly "reliable" answer; but if he is asked about an item which he never had to choose before, he cannot give a meaningful answer—he literally does not know how much he would pay for it. Thus, hypothetical choices induce inaccurate responses, and phenomena such as anchoring and starting point bias arise.

The crucial issue is one of degree: When we ask a hypothetical valuation question, relatively how much signal does the response contain and how much noise? If the response is entirely noise—people just do not know how much they would pay to save the gray whale and they give entirely meaningless answers—then the results of CV surveys are useless. If the response contains some signal, as well as noise, then we can perhaps obtain useful information from CV surveys. Suppose we be lieve that CV responses do contain some signal. Then the question arises: How can we decode the signal? Do we have a theory of how the noise is generated *which would enable us to unscramble the response and uncover the signal embedded in it?* I will give some examples of a decoding procedure in the next section; but, first, I want to emphasize the significance of the underlined clause above. It is not enough to have psychological experiments which reveal that respondents fall prey to starting point bias or otherwise give inaccurate responses to hypothetical choices. We need a psychological model of response formation in these circumstances—in effect, we need a formula for the bias—if we are to derive useful information from the responses.

Ultimately, the answer to these questions must come from controlled laboratory experiments, such as those pioneered by Vernon Smith. Very little of this work is germane to the issues at hand. This is because, as I understand it, in most experiments the organizer induced a well-defined utility function in the participants. Thus, the participants knew their preferences exactly, and the issue was how they acted on those preferences in various market settings. What is needed, instead, is to somehow mimic the type of uncertainty about one's own preferences that I believe arises in real-world CV studies and then examine how participants respond to different types of interview format (e.g., iterative bidding with starting points, noniterated bidding with payment cards, etc.). Moreover, one needs to vary the degree of the participants' uncertainty about their own "preferences" and see at what point phenomena, such as starting point bias, begin to emerge. In the near future, Richard Carson and I plan to conduct some experiments on these lines. In the meanwhile, I want to focus on some other procedures, essentially statistical techniques, for deriving useful information from CV data.

In the next section I discuss some procedures for decoding the starting point "noise" in iterative bidding, continuous response CV experiments, and this is followed by a discussion of the role of discrete response experiments as a possible vehicle for obtaining more reliable valuations.

Decoding Starting Point Bias in Iterative Bidding

The first decoding procedure that I will discuss was proposed by Thayer (1981); I want to motivate it slightly differently and suggest a new method of implementing it. The basic notion is that there are *two* separate models—a model of the individual's "true" preferences and a model of the individual's behavior in response to the CV survey. The first model is a conventional utility-theoretic model of compensating or equivalent surplus. An individual derives utility from money income, y, and from an environmental good, q; in addition, his preferences may vary systematically with either observable attributes (age, sex, previous experience of the environment, etc.) which are represented by the vector t. The resulting utility function will be written u = v(q, y; t).[1] Suppose that the individual's current income is y_0 and the current supply of the environmental good is q_0. The individual is going to be asked about his willingness to pay for an improvement in the supply of the environmental good to a new level qi > q_0; his true willingness to pay is the quantity C which satisfies

(1) $$v(q_1, y_0 - C; t) = v(q_0, y_0; t).$$

Although C is the individual's true willingness to pay, as revealed by introspection in a moment of quiet reflection, it is not necessarily the amount that he gives to the interviewer in the CV study. Two stories can be told here: the individual becomes bored during the iterative bidding procedure and seeks to limit its duration. Alternatively, when he receives a starting bid, he takes this as an indication of what a reasonable valuation *ought* to be (i.e., anchoring) and wonders whether he should revise his own bid in the light of this "information." Both stories may lead to the same model of response behavior: given his true willingness to pay, C, and the starting bid, C_s, the individual selects his response—his final bid—C_f by solving

(2) $$\underset{C_r}{\text{Minimize}}\ a(C_f - C_s)^2 + b(C - C_f)^2$$

where a $\geq$ 0 and b $\geq$ 0 are behavioral parameters. Under the first scenario, in solving (2) he is trading off a desire to be truthful against his impatience with the iterative bidding format; under the second scenario, he is deciding how much weight to place on his original, instinctive valuation as against the new "information." In either case, the behavioral response model (2) has the solution

(3) $$C_f = (1 - \delta)\, C + \delta C_S$$

where δ = a/(a + b); and 0 $\leq$ δ = δ $\leq$ 1.

Thayer and others have applied this model by fitting a regression equation of the form

(4) $$C_f = \ y + \delta C_S$$

and then deriving C from the relation

(5) $$C =$$

using the estimates of *y* and 8 in (4). However, this assumes the validity of the behavioral response model (2) and does not subject it to a statistical test. The distinctive feature of the model (2) is that the final bid, C/, is a weighted sum of the starting bid, C_2, and the individual's true valuation, C, where the weights are nonnegative and sum to unity.[2] How can this prediction be tested? By postulating a specific .functional form for the utility function v(q, y; t); solving (1) for the true valuation function C = C(q_0, qi, y<>; t); fitting regression model

(6) $$C_r = yC(q_0, qi, y_0; t) + SC_S;$$

and testing the hypotheses that *y* = (1 — 8) and Os 8 s I.[3] For example, suppose one hypothesizes that utility is a monotone transformation of

(7) $$v(q, y; t) = [y^* + 0ft)qT" \ \mathit{0} > 0$$

where 0(t) can be some parametric function of the individual characteristics, t; this is a constant elasticity substitution (CES) function with an elasticity of substitution between money and the environmental good of cr = (1 - p)"^{1}. Then, (3) becomes[4]

(8) $$C_f - (1 - 8)\{y_0 - [0(qo^p - q/) + yo^p]^{1/p}> + 8C_S.$$

It would, of course, require nonlinear least squares to estimate (8) but such computer programs are readily available.

[1] This may be a direct utility function or an indirect utility function which arises after the individual has optimized with respect to all private (nonenvironmental) goods.

[2] To be sure, if one estimates (4), he can test whether 0 ss Ss 1; but he cannot test the other implication of (3), namely that the coefficients (1 — S) and 8 sum to unity.

[3] Rowe et al. propose a willingness to pay function for the final bids resulting from iterative bidding surveys that takes the form of (6), but no constraint on the sum *(y* + 8).

[4] This example assumes that the supply of the environmental good can be measured along some continuous scale. An alterna tive scenario is where there are only two states representing, respectively, the presence or absence of the environmental good. In this case, q is effectively a binary-valued variable. Here one could equivalently write the utility function as v(0, y; t) = [y° + 6UO] and v(l, y; t) = [y° + 0,(t)], from which one obtains

$$C_p = (1 - S)[y_0 \mathbf{- (ft - ft + yo")"*]\ \ + SC_S-}$$

This procedure yields estimates of the parameters of the utility function, *0* and *p,* as well as the parameter of the behavioral response model, 8; it also permits us to test the behavioral response model against an alternative which I outline below. The test of the validity of the behavioral response model is conditional on the maintained hypothesis of the particular utility function, and one ought to repeat it using several different utility functions for the sake of robustness.

It is instructive to compare the nonlinear regression model (6) or (7) with the more conventional linear regression model (4). If one assumes that (8) is the correct behavioral relation, it follows that (4) is misspecified since it contains a constant term and omits the variables q_0, qi, and y_{fl}. More generally, (4) implies that the true willingness to pay is constant across the sample and is independent of income, the attributes t, and any random factors which affect preferences. This is most implausible; from casual inspection of survey data, one often obtains an impression of striking differences in the valuation of environmental goods. If the formula for C were linear in the coefficients, of the general form C = X/3, one could apply Their s specification analysis to compare (4) with (6). In that case, making the reasonable assumption that the starting bids C_s are uncorrelated with the regressions in X, it would follow the estimate of S from (4) is *unbiased* and the estimate of the constant ternu'y.measuresU — 8)X/3,whereXisthemean value of the regressors across the sample; thus, (5) produces an estimate of C, the average true willingness to pay across the sample. However, the formula for C will usually be nonlinear in the coefficients, as illustrated by (8), and Their s specificationanalysiscannotbedirectly applied. In this case itisunclearwhetherfitting(4) instead of (8) yields a biased estimate of 8.

So far, I have said nothing about the disturbance term in these regression models: I have implicitly assumed that a random error is added to the right-hand side of (4) or (8). This is appropriate if one assumes that there are measurement errors in recording the final bids, Cf, or that the presence of unobservable variables and/or variation in preferences justifies the imposition of an *additive* error on the formula for C(q_0, qi, y; t). An alternative approach is to introduce an explicit random element directly into the utility function v(q, y; t)—for example, in the context of (7) one might assume that 0(t) is, say, a lognormal random variable with some mean and variance which depend on t. This is the so-called "random utility" approach which will be discussed further in the next section. In effect, this approach converts (8) into a random coefficient regression model. Because of the nonlinearity, it would need to be estimated by maximum likelihood rather than weighted least squares. Letting f_e(*) be the density of *0,* the density of C_f can be obtained from f_e(-) by change of variables; in particular,

(9) Pr{C, = X} =

f L - (X-gCJ1P_vp1
y° (1- S) yo .
"T-1—<s^?5—/

I now want to discuss another model of individual behavior in responding to CV surveys as an alternative to (2); this is prompted by the recent papers of Carson, Casterline, and Mitchell (1984). Suppose that, given his original willingness to pay, C, the individual engages in a special form of yea saying during the course of the CV interview: if C_s > C, he raises his valuation of the environmental good *beyond* C_s; and if C > C_s, he lowers his valuation *below* C_s. Specifically, suppose that

(m c _ C_s + iMC, - C), if C_s > C ^
} f c, - *(C - C_s), if C_s< C

where 0 < *$* == 1. Then

(11) C_F = (1 - 8)C + SC_s

where 6 = (1 + i//) ^ 1 and (1 - S) = -t//, with — i s (1 — 8)^0. The behavioral response function (11) is formally similar to (3), but the sign restictions on the coefficients of C and C_s are quite different. As before, one would substitute some specific parametric true valuation function, such as (7), into (11); and one could either assume an additive disturbance term and apply nonlinear least squares or adopt the random coefficient formulation and use maximum likelihood. One further point about (10) should be noted; when C_s < C, there is probably a constraint that C_f ^ 0. Thus, when C_s < C, the response might be C = max[0, C_s — i//(C — C_3) in which case (11) becomes

(11") Q = max [0, (1 + 8)C + 8CJ.

which is a nonlinear Tobit model.

Observe that, whereas the behavioral model (2) implies that C_f ^ C_s« C ^ C_s, the behavioral model (10) implies that C_f ^ C_s •» C ^ C_s. A variant of each model can be derived by following the suggestion of Carson, Casterline, and Mitchell that the behavioral response may vary according to $C > C_s$ or $C < C_s$. Thus, the loss function in (2) becomes

(2') minimize
Cj
fa(C_F - C_s)^2 + bi(C - C_f)^2, if C> C_s la(C, - C_s)^2 + b_2(C - C_f)^2, if C < C_s

which leads to

m c = (1 ~ 6l)C + SlCs, tf Cf > Cs
f (1 - 6s)C + &>C_S, if $C_f < C_s$

where 0^ Si s 1, i = 1,2, which is a switching regression model. Similarly, (10) becomes

ClO'1 f — s ~*~ T'l'^"S C), it O_s > C-
f **c_s + <MC - c_s), ifc$_s$<c**

which leads to

U1 ,,, c f ^ (1 - SOC + 8iC_a, if $C_f > C_s$
max [0, (1 - 52)C + *§,Cs]*
if $C_f < C_s$

where Si > land-1 < (1 - 6i)^0,i= 1,2, which is a mixed switching regression/Tobit model.

Discrete Response Surveys

The models (3), (30, (10) and (100 are examples of one method of dealing with the problem of hypothetical bias in CV surveys, viz., to assume that the bias takes a specific form (anchoring on the starting point) and then to decode it. However, some features of this approach may still raise doubts. In particular, it presumes that the individual has a precise initial valuation of the environmental good, which he adjusts in response to the starting bid presented in the CV survey. This contradicts my earlier assertion that individuals do not know their true preferences until they actually make a real choice—as opposed to a hypothetical one. Perhaps it is unwise, whether using an iterative or some noniterative procedure, to hope to pin people down to *exact* values of their willingness to pay for hypothetical changes in the supply of environmental goods. I want to suggest that their responses will be more reliable if they are required only to place bounds on their willingness to pay. Perhaps one can never give a meaningful response to the question: "How much are you willing to pay?"; but one *can* give a meaningful response to such questions as, "Are you willing to pay at least $5.00?" or "Is your willingness to pay less than $20?" This leads me to argue that certain surveys involving only discrete responses are inherently more reliable than the conventional surveys which require a continuous response.

The first example of a discrete-response experiment is the study by Bishop and Heberlein, referred to earlier, in which a specific amount of money was designated for each subject; and the subject was asked if he would be willing to pay that amount for the environmental good. The resulting responses of "yes" or "no" were correlated with the amount of the offer using logit analysis. In Hanemann I have discussed how the statistical logit model can be related to an underlying utility theoretic model of individual behavior. The essential notion is the concept of random utility mentioned in the previous section, i.e., the notion that, from the viewpoint of the econometric investigator, the individual's utility function contains some stochastic components which are modeled explicitly. The presence of these components can be explained in various ways. They can be taken as representing errors of measurement or unobserved variables which are known to the individual but not the econometric investigator; alternatively, they can be taken as representing random variation in preferences across individuals. The random terms will be denoted e, which can be a scalar or vector; and the utility function will be written: u = v(q, y; t, e). Accordingly, if the individual is confronted with the amount $A and agrees to pay it, the probability that this response occurs is

(12) Pr{willing to pay A}
= Pr}v(q!, y_0 - A; t, c)
^ v(q_0, y_0; t, e)};

if the individual is unwilling to pay $A, the probability of that response is one minus the probability in (12). Depending on the distribution of e and the functional form of v(-), these probability statements may correspond to a logit or probit statistical model although possibly one that is nonlinear in the parameters. To continue the example based on (8) in which *0* is treated as a random parameter (i.e., *6* = e), the probability statement in (12) becomes

(13) Pr{willing to pay A}
= Prf_e < (y°-^{A})^P]-
H ~ (Qo" - «,') I'

if *6* is lognormal, this is a nonlinear probit model.

I have argued that the individual does not know very accurately what is the most that he would be willing to pay for a change in the environmental good, but he *does* know reasonably well whether it is less or more than \$A. Nevertheless, in conducting benefit cost analysis, we *do* actually need to estimate (or infer) the most that the individual would be willing to pay. Suppose that he is asked, a la Bishop-Heberlein, whether he would be willing to pay \$8.00 for an improvement in the supply of the environmental good and he answers, "yes." Then, we know that \$8.00 is a lower bound on his true willingness to pay, but we have *no* idea what the upper bound is; if the cost of the change amounts to more than \$8.00, this is a serious problem. In Hanemann I have shown how one can derive, from the responses to Bishop-Heberlein questions, estimates of the maximum willingness to pay for an individual with given income, y_0, and characteristics, t. The key to this procedure is to postulate a specific, parametric random utility model for the individual, set up the resulting statistical model for the responses to the survey as in (12) or (13), fit the statistical model (using the observed responses thereby recovering an estimate of the coefficient of the utility model, and then use the estimated utility model to calculate willingness to pay as in (1). However, in the random utility context, there is the complication that the willingness to pay is a random variable. The analog of (1) is

(10 $\quad v(qi, y_0 - C; t, e) = v(q_0, y_0; t, e)$

which implies that $C = C(q_0, qi, y_0; t, e)$. Two plausible procedures are to use the mean or median of the distribution of $C(q_0, qi, y; t, e)$ as our estimate of the individual's willingness to pay; I denote these by C^+ and C^*, respectively. In Hanemann I show how C^+ and C^* can be calculated from the fitted statistical response model; in terms of the graph in Figure 1, C^+ corresponds to the shaded area below the graph while C^* is the value at which the estimated response probability is 50 percent.

The logic of this welfare calculation can be illustrated with a simplified example. Suppose that everybody has *exactly* the same preferences. The utility function is nonstochastic (no random variation in tastes) and is identical for everyone (the shift variables, t, and income, y, are identical). In particular, suppose that the

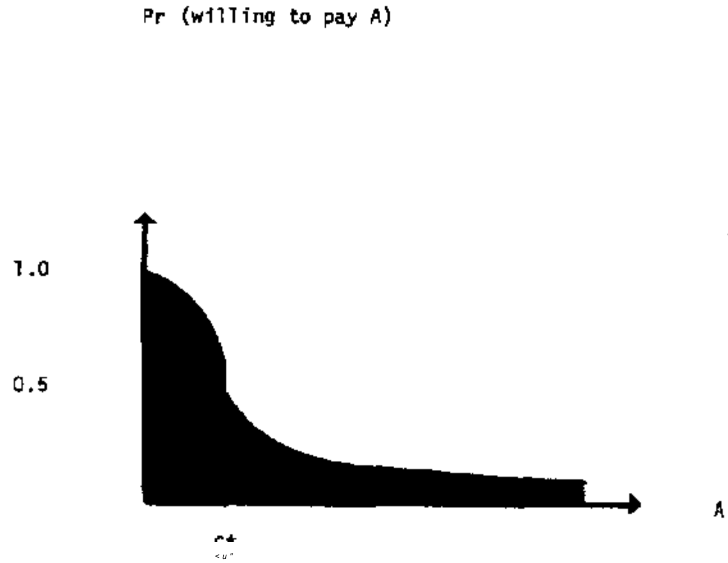

Figure 1. Response probability function and welfare measures in Bishop-Heberlein survey.

common willingness to pay for the change in the environmental good is \$10. I go out and interview a sample of people from this population. I do not ask them directly how much they would be willing to pay; instead, as in Bishop and Heberlein, I name a particular sum, \$A, and ask whether they would be willing to pay this amount. I find that, when A < \$10, everybody answers, "yes;" but when A > \$10, everybody says, "no." I infer that the maximum willingness to pay is \$10. It will be seen from the graph of the response function in Figure 2 that \$10 corresponds to both the wel-

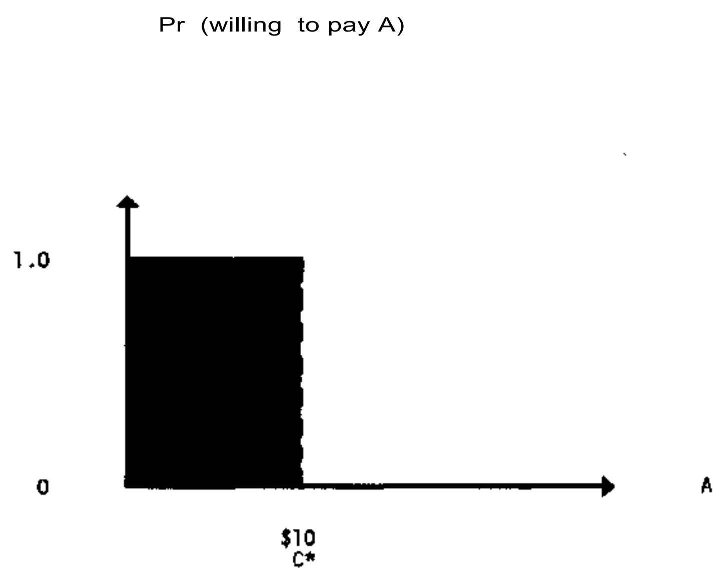

Figure 2. Response probability function and

fare measures, C^+, and C^*. Thus, although this is a fanciful example, it serves to motivate these welfare measures. In the real world, of course, everybody's preferences are *not* the same, for two reasons: systematic differences in preferences associated with variation in the observed shift variables, t, and random differences in tastes associated with e. The first type of variation in preferences does not affect my argument. I could draw a different response function, as in figure 2, *for each* possible value oft. The random variation in preferences associated with e does alter the story. Because of this, the response function has the shape shown in figure 1 rather than in figure 2 and the two measures, C^+ and C^*, no longer coincide in general, but the general logic underlying the two figures is the same.

One can characterize the Bishop-Heberlein type of survey as giving a single piece of information about the individual's willingness to pay for the change in environmental quality—either a lower bound (if the individual's response is, "yes") or an upper bound (if the response is, "no"). Clearly, one could go beyond the Bishop-Heberlein question while remaining within a discrete-response framework by developing *both* a lower *and* an upper bound on the maximum willingness to pay. If the individual answers, "yes," to the initial question, choose a larger amount, A', and repeat the question; if he answers, "no," to paying A', then you know that A and A' are, respectively, lower and upper bounds on his willingness to pay. Instead of (12), the corresponding probability statement for the observed response becomes

n/n pj willingness to pay 1
U4J ^(bounded by A and A'j
Pr{v(q$_1$, y$_0$ - A; t, e)
s= v(q$_0$, y$_0$; t, c)
and v(q$_{15}$ y$_0$ - A'; t, e)
^ v(q$_0$, y$_0$; t, «)}.

In the case of the CES utlity model, for example, one obtains

n <*n* PT! willin$_{\text{s}}$ness to P^{a}V 1 =
V ' 1 bounded by A and A' J

- Pr f (yo " Ar 1 qo^p -
qx^p

^ $_e$^ (Yo - *Ay }*
qo^p ~ q/ I'

Thus, while remaining within a discrete-response framework, one may actually obtain two pieces of information about the individual's preferences. Now, this may seem suspiciously like the iterative bidding, continuous response survey that I criticized earlier. The difference is one of degree: How finely does one attempt to bracket the individual's preferences? If the bracketing is very fine, one does, indeed, generate something that is equivalent to a continuous response model. However, what I have in mind is to keep the bracketing fairly coarse by offering the individual several broad ranges of money values and asking him which range contains his valuation of the change. In this way, one avoids trying to do what I believe is psychologically impossible—pinning the individual down to an exact statement of his valuation of the change.

To summarize, there is a trade-off between the size of the sample in the survey and the quantity of information obtained from any respondent, on the one hand, and the underlying reliability of the responses in the light of psychological obstacles to valuing hypothetical choices, on the other hand. There is no doubt that a pure continuous response experiment yields the most information in the statistical sense or that a finely bracketed, discrete-response experiment yields more information than either a coarsely bracketed, discrete-response experiment or a Bishop-Heberlein discrete-response experiment. However, as one passes from the Bishop-Heberlein experiment to a pure continuous response experiment, in my opinion, the danger of obtaining unreliable and psychologically meaningless response increases. The optimal point along this continuum can be determined only by conducting both Monte Carlo simulation studies, to measure the change in statistical information , and controlled laboratory experiments, to measure the change in reliability. I hope to conduct such studies in the near future.

Conclusions

In this section I want to comment on two topics: the role of explicit, parametric utility models and the possibility of conducting CV-like studies aimed toward eliciting not forecasts of values but, rather, forecasts of behavior.

Throughout this paper, I have emphasized the adoption of explicit parametric utility functions, such as the CES function (7), in order to formulate statistical models for analyzing the responses to CV surveys—be they statistical models of continuous response surveys, such as (8), (9), (10), (11), (100, or (11"), or statistical models of discrete-response surveys, such as (13) and (15). I should point out that this can be extremely dangerous if it is not done properly. Once a function, such as the CES, is adopted, it becomes *a. maintained* hypothesis; we allow the data to tell us the correct value of *p* and *0,* but we may forget to inquire of the data whether the CES form, as a whole, is correct. This can be avoided only by employing *a. variety* of parametric utility functions and then employing nonnested hypothesis tests to see which is the correct form and/or examining the robustness of valuation estimates across different function specifications. Another factor that should be borne in mind is the desirability of employing random coefficient versions of the random utility hypothesis exemplified by randomness of *B* in the CES model as opposed to the conventional, additive-error random utility formulations that appear almost everywhere in the literature on logit and probit. I strongly suspect that the additive-error formulation induces too little variation in preferences (or induces the *wrong kind* of variation in preferences) and reduces the goodness of fit. If these steps are not taken, we may obtain results which are essentially meaningless because of the straight)acket of an inappropriately specified utility model.

Assuming that this warning is heeded, there are at least three reasons why I believe that one ought to work with explicit parametric utility functions. The first reason is to insure consistency in modeling the individual's responses; this is the analog of the integrability conditions in conventional demand theory. In Hanemann I have pointed out a violation of these consistency conditions in the ad hoc statistical logit model employed by Bishop and Heberlein. The second reason is that we often need to be able to extrapolate from the sample of individuals in the survey to others in the population who have quite different incomes or socioeconomic attributes (t); similarly, we may want to extrapolate to a different environmental change from that employed in the survey questions. Both of these are impossible unless we have an explicit utility model. The third reason has to do with the quite common observation that it is often more difficult to obtain reliable responses to willingness-to-sell, as opposed to willingness-to-pay, questions in contingent valuation surveys. If we limit ourselves to willingness-to-pay questions but employ an explicit utility model, we can readily compute what the willingness to sell would be from the fitted utility model.

Lastly, I want to point out that, although most CV studies (whether of the continuous- or discrete-response variety) have sought to elicit expressions of the valuation of hypothetical changes, there is no reason why one cannot use similar procedures to elicit forecasts of *behavior* in hypothetical circumstances. Moreover, there may be an advantage in eliciting forecasts of behavior because they may be less hypothetical than forecasts of valuations. To give a concrete example, suppose we want to measure the welfare loss to users of shutting down a recreation site. If we had travel cost data, we would postulate a utility function, derive and estimate the demand function, and then calculate some measure of consumer's surplus. To keep the example simple, suppose we postulate the no-income-effects demand function

(16) x = a — bp + «

where the disturbance term, e, is treated as coming from a random utility formulation. Then, the loss of consumer's surplus for an individual whose travel cost is p_0 is

<"> c -1(^-4

If, however, we conduct a CV study, instead of asking individuals how much they would be willing to pay to avoid the closing of the site, we could, instead, ask them (for example) how high the price would have to rise before they ceased visiting the site completely. If we did this, it might still be possible to recover their preferences from their responses to this question and employ the fitted demand (utility) model to calculate their welfare loss. Our question could be of the discrete- or continuous-response variety. Following Bishop and Heberlein, we could ask them: "If the price rose by $A per visit, would you cease visiting the site completely, yes or no?" Suppose an individual answers, "yes." The probability of obtaining this response is

$$(18)\quad \Pr\begin{Bmatrix}\text{cease visiting if}\\ \text{price rises by } A\end{Bmatrix} = \Pr\{\alpha - \beta(p_0 + A) + \varepsilon \le 0\}$$
$$= \Pr\{\varepsilon \le -\alpha + \beta(p_0 + A)\}$$

which is a simple logit (or probit) model. Together with Ivar Strand, I am currently conducting an experiment based on this type of discrete-response CV data as well as the conventional travel cost approach to fitting demand functions; it remains to be seen whether the demand (utility) model implied by the CV responses is consistent with the demand function fitted to the data on actual site visitation behavior.

References

Bishop, R. C., and T. A. Heberlein. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased." *American Journal of Agricultural Economics* 61(1979):926-30.

Boyle, K. J., R. C. Bishop, and M. P. Welsh. "Is Starting Point Bias a Problem in the Iterative Bidding Format of Contingent Valuation Studies." *Land Economics* (forthcoming). ,

Carson, R. T., G. L. Casteriine, and R. C. Mitchell. "A Note on Testing and Correcting for Starting Point Bias in Contingent Valuation Surveys." Discussion Paper D-116, Quality of the Environment Division, Resources for the Future, January 1984.

Hanemann, W. M. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics* 66(1984):332-41.

Rowe, R. D., R. C. d'Arge, and D. S. Brookshire. "An Experiment in the Value of Visibility." *Journal of Environmental Economics and Management 7* (1980)11-19.

Thayer, M. A. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence." *Journal of Environmental Economics and Management* 8(1981):27~43.