Environ Resource Econ (2007) 37:181–210
DOI 10.1007/s10640-007-9124-5

Incentive and informational properties of preference questions

Richard T. Carson · Theodore Groves

Published online: 16 May 2007

Abstract Surveys are frequently used by businesses and governments to elicit information about the public's preferences. They have become the most common way to gather preference information regarding goods, that are not (or are not yet) bought or sold in markets. In this paper we apply the standard neoclassical economic framework to generate predictions about how rational agents would answer such survey questions, which in turn implies how such survey data should be interpreted. In some situations, the standard economic model would be expected to have no predictive power. For situations where it does have predictive power, we compare different survey formats with respect to: (a) the information that the question itself reveals to the respondent, (b) the strategic incentives the respondent faces in answering the question, and (c) the information revealed by the respondent's answer.

Keywords Contingent valuation · Stated preference surveys · Incentive compatibility

1 Introduction

Businesses and governments frequently use surveys to help determine the relevant public's preferences toward the different products they might offer, or different policies they might adopt. Applications are particularly common in environmental valuation (Mitchell and Carson 1989; Bateman et al. 2002; Pearce et al. 2006) in both developed (Pearce and Markandya 1989; Pearce 2006) and developing countries (Pearce et al. 2002), health care (McDowell and Newell 1996), marketing (Louviere 1994), political science (Lavrakas and Traugott 2000)

Earlier versions of this paper have been presented as a plenary address to the European Association of Environmental and Resource Economics in Oslo, as an invited paper at the Japanese Forum on Environmental Valuation meeting in Kobe, and at a NOAA conference on stated preference methods.

R. T. Carson (✉)· T. Groves
Department of Economics, 0508,
University of California, San Diego,
La Jolla, CA 92093, USA
e-mail: rearson@ucsd.edu

and transportation (Hensher 1994). As long as the economic agents (hereafter, *agents*) being surveyed believe that their responses might influence the actions taken by businesses or governments (hereafter, *agency*), the standard economic model suggests that agents should respond to the survey in such a way as to maximize their expected welfare.

Given the billions of dollars spent annually on surveying, and the frequently voiced concerns that marketing surveys determine the fate of products and that major political decisions are largely poll-driven, the position of many economists that survey responses should be ignored as a source of preference information is somewhat surprising. These economists seem to regard survey responses as either completely meaningless because they are answers to hypothetical questions or else as completely useless because agents will respond strategically. The first argument violates the standard rationality condition assumed of agents, if in fact agents believe that agency decisions are being made at least in part on the basis of their survey responses. The second argument stops short of the more relevant question, namely "What are the agents' strategic incentives and how should those incentives influence their responses?"

In this paper we explore the implications of the economic maximization hypothesis for the behavior of rational agents answering preference surveys.[1] The literature on neoclassical choice theory and mechanism design (Hurwicz 1986; Groves et al. 1987; Varian 1992) provides the theoretical foundation for our work. This body of work can be contrasted with other psychologically-based theories (e.g., Kahneman et al. 1982).[2] We believe that at least some of the evidence put forward in support of those theories, particularly with respect to what differences should be expected when asking questions with different response modes, has been incorrectly interpreted.[3] In the model informally presented here, agents are assumed to consider whether the aspects of the described scenario are plausible and how the survey results are likely to be used; in particular, how the outcome might be influenced by their responses. These assumptions, combined with the basic maximization hypothesis, are capable of yielding a surprisingly rich picture of the manner in which agents would respond to survey questions.

A major reason why many economists view survey-based estimates of economic values with suspicion is a body of empirical results which seem inconsistent with economic intuition. These anomalous results have often been interpreted as evidence of (a) the hypothetical nature of the question, (b) strategic behavior,[4] or (c) preferences which are either ill-defined or inconsistent with economic theory. In attempting to systematically categorize these anom-

[1] Note that in this paper we only consider survey questions concerning agents' preferences and not questions concerning personal characteristics such as age and income.

[2] See McFadden (1999), Rabin (1998) and Sugden (1999) for thoughtful surveys of these issues by economists.

[3] It is important to note that we are not denying that some of the phenomena identified in the psychological literature might exist. Rather, we are taking the position that the implications of the standard neoclassical approach should be fully developed before concluding that a particular result is inconsistent with standard theory.

[4] The possibility of strategic misrepresentation of preferences has long been seen as a problem in public economics. Samuelson (1954) argued "It is in the selfish interest of each person to give false signals to pretend to have less interest in a given collective activity than he really has" and made specific reference to the possibility of strategic behavior in surveys. Samuelson's admonition, repeated in many textbook discussions of public goods, had a profound effect on how many economists view survey questions. The mistaken inference made from this admonition was to equate strategic behavior with lying. As the term is used in modern mechanism design theory, strategic behavior is merely synonymous with rational agents maximizing their self-interest. Mechanism design theory has shown that, in some instances, the optimal strategic behavior for agents is in fact to truthfully reveal their preferences. Whether this is the case or not depends upon the particular structure of the preference question.

alies, it becomes apparent that there is an antecedent question: Does a survey question need to meet certain conditions before it can be expected to produce useful information about an agent's preferences?

This question is easy to address. First, the agent answering a preference survey question must view their responses as potentially influencing the agency's actions.[5] Second, the agent needs to care about what the outcomes of those actions might be.[6] We will term survey questions that meet these two criteria as *consequential*, and those that don't as *inconsequential*. That is:

> *Consequential survey questions:* If a survey's results are seen by the agent as potentially influencing an agency's actions and the agent cares about the outcomes of those actions, the agent should treat the survey questions as an opportunity to influence those actions. In such a case, standard economic theory applies and the response to the question should be interpretable using mechanism design theory concerning incentive structures.

> *Inconsequential survey questions:* If a survey's results are not seen as having any influence on an agency's actions or the agent is indifferent to all possible outcomes of the agency's actions, then all possible responses by the agent will be perceived as having the same influence on the agent's welfare. In such a case, economic theory makes no predictions.

Most preference survey questions asked by businesses and governments meet the two criteria for being consequential, and hence, can be interpreted in economic terms.[7] There are, however, many preference survey questions which do not. While a lot of these inconsequential survey questions could be characterized as issuing from laboratory exercises with under-graduates, there are plenty of real world examples.[8] It is pointless to try to explain apparent economic anomalies in inconsequential survey questions, since all responses to such questions have the same effect on the agent's welfare. We thus formally reject the notion, sometimes advanced by proponents of preference surveys, that when a respondent perceives no gain or loss from how a preference survey is answered, the respondent always answers truthfully. While such an assumption may be true, there is no basis in economic theory to either support or deny it.

For consequential questions, we examine four key issues which illustrate both the power and the limitations of economic theory to interpret a large body of empirical evidence about

[5] Carson et al. (2004) explore the issue of probabilistically versus deterministically influencing a decision which is crucial to both the use of surveys and many economic experiments and show that neoclassical theory is applicable unless the influence is zero.

[6] For instance, a non-smoker may not care about the addition of a new type of cigarette with a much lower nicotine level and a higher price than currently available cigarettes. Confusion often exists over the magnitude of the possible change in utility from agency's action and the incentives the agent faces in the response given to the question. The magnitude of the utility change generally does not influence the incentive structure of the question as long as there are nonzero differences in utility levels resulting from different agency actions. The magnitude of the utility change can, however, influence agents' participation in the survey.

[7] Marketing research firms, in particular, face a constant battle between asking questions to only those who are currently using a product category versus trying to reach the larger and harder-to-identify population of all potential users. For public goods provided via taxation the situation is generally easier: Even if a respondent does not care whether the good is provided at zero cost, he or she will care about its provision if the tax cost is positive.

[8] Inconsequential preference questions can most often be identified by having one or more of the following characteristics: (a) being asked of a population or at a location that is irrelevant from the perspective of an agency seeking input on a decision, (b) providing few, if any, details about the goods and how they would actually be provided, (c) asking about goods that are implausible to provide, or (d) about an implausible prices for them.

such questions. First, we look at the properties of *binary discrete choice questions* under different circumstances. In particular, we examine whether such question formats are *incentive compatible,* in the sense that a truthful response to the actual question asked constitutes an optimal strategy for the agent.[9] The empirical evidence suggests that such questions often work well: they predict actual behavior quite closely and are sensitive to factors such as the scope (quantity or other attributes) of the good being valued. Note, however, that there are also instances where such questions perform quite poorly. Second, we consider reasons why responses to *repeated* binary discrete choice questions (e.g., double-bounded dichotomous choice) by the same respondent are often inconsistent with each other. Third, we consider complications introduced when attempting to value *multiple* goods, first by a sequence of pair-wise comparisons, and then by the increasingly popular multinomial choice experiment format. Fourth, we look at whether binary discrete choice questions and *open-ended continuous-response questions* should produce similar estimates of parameters such as mean or median willingness-to-pay (WTP). When doing so, we shall pay particular attention to the issue of what role, if any, information on cost might have on reported WTP values. Before examining these issues, we discuss what is called the *face-value property*.

2 The face-value property

Economists tend either to reject preference survey results out of hand or treat the answers as truthful responses to the questions asked. We term the latter position as taking survey answers at *face value*. Many who reject the use of surveys do so because the results are anomalous if taken in that manner. However, taking survey answers at face value is likely to be wrong in many circumstances, even when the two criteria for consequential survey questions are met.

The face-value property is the behavioral property that respondents always *truthfully* answer the *specific* survey question being asked. There are two aspects of this property: (a) that respondents always answer truthfully, and (b) that respondents always correctly understand and answer the question being asked. While the mainstream economic position is that property (a) is dubious due to strategic behavior, it is routinely assumed in marketing research, political polling, psychology, sociology and other fields heavily dependent upon survey research. In contrast, while economists who use survey results routinely seem to believe (b), survey researchers have shown this to be a dubious assumption (Sudman et al. 1996).

Interpreting responses to survey questions appropriately requires consideration of when one or both components of the face-value property might fail, and how responses should be interpreted when this happens. Even surveys that give agents an incentive to misrepresent their preferences can yield useful information, and some survey formats may be expected to induce different types and degrees of misrepresentation than others.[10]

[9] Note that this definition of incentive compatibility makes explicit the assumption, left unstated in the mechanism design literature, that agents correctly interpret the question, which may not happen with poorly worded surveys.

[10] That is to say, under some question formats, the expected direction of the bias in responses can be theoretically predicted, and in some instances empirically confirmed. In such cases, the survey's results, even if biased, may be useful and often sufficient for agency decision-making (Hoehn and Randall 1987). Some studies have directly manipulated the incentives for preference revelation. An early example of such a study was Cronin (1982), who looked at WTP for improving water quality on the Potomac River. Cronin found that the WTP from the subsample of agents who were given a statement to the effect that the Federal government was likely to pay for most of the cost of the project was substantially higher than the subsample not given this

The survey research community's usual rationale for the possibility that respondents may answer a different question than the one being asked is simply that respondents may not understand the question actually asked, and instead, answer the question they *think* is being asked. Part of the survey designer's art lies in crafting language that elicits the answer to the question the researcher intends to ask (Payne 1951). This issue needs to be taken particularly seriously for survey questions regarding non-marketed goods or new consumer products, and the development of questionnaires describing such goods is among the more difficult of survey design tasks. If survey responses are to be taken at face value, the question as written should elicit the answer to the question intended by the designer. If this does not happen, the results can easily be viewed as implying violations of economic theory, when what has really happened is that quite simple: the agents have answered a different question.

A further issue concerns preference questions with implausible premises. A common example includes asking a choice question involving an implausibly high or low cost for providing the good. In such cases, respondents are likely to substitute what they consider to be a more realistic cost, and answer on that basis. This can result in violations of the predicted weak monotonicity of the response to changes in cost at very low or very high costs.[11] Another common example occurs when the agent is told that the answers to the survey questions will not influence agency decisions even though it is obvious that the information is be collected at considerable expense. The agent, of course, should ignore this language and answer the questions consistent with how the agent thinks the agency will use the information.

A different variant of this issue arises when the feasibility of the agency actually being able to deliver the good is in question. This can happen, for instance, a claim that a proposed risk reduction program would be 100% successful, which is likely to be discounted by agents. It can also produce the appearance of insensitivity to scope of the good being valued since a "large" variant of the good will typically be seen more likely to be supplied than a "small" variant of the good and an agent should take into account the probability of provision in their choice behavior.[12]

Credibility problems can also occur when a respondent is given inconsistent information at various points in a survey. Examples include providing two different cost numbers in a double-bounded dichotomous-choice elicitation format, or asking respondents about the provision of two different levels of the same public good at different points in a survey. Further, there are limits to the range of preference questions that a respondent will informatively answer. Although survey questions can extend the range of goods and their attributes (including price) considerably beyond what agents have previously experienced, any counterfactual scenarios must be credible portraits of possible future outcomes. It may also be the case that scenario elements that work well for most agents will fail to have their assumed

Footnote 10 Continued
statement on possible cost sharing. Lunander (1998) and Posavac (1998) also provide examples of directly manipulating the incentive structure of preference questions, again with the result that the incentives work in the expected direction.

[11] The often observed practice of asking agents if they want the good if were free or cost only a very small amount may be problematic because an agent should not answer this question with a zero or trivial cost in mind since they will have to pay for the good if provided and it cannot be provided at zero cost. This may explain why there is such a large fraction of the sample that typically indicates they do not want the good even though it seems desirable to have if the cost of provision really was free.

[12] Tests of scope sensitivity in surveys accentuate this problem by using as nearly identical language as possible while it make take a much more elaborate description of how the larger good will be provided to make the perceived probability of provision of the two goods equivalent. In extreme cases, it may be impossible to convince agents that a very large good can be credibly provided.

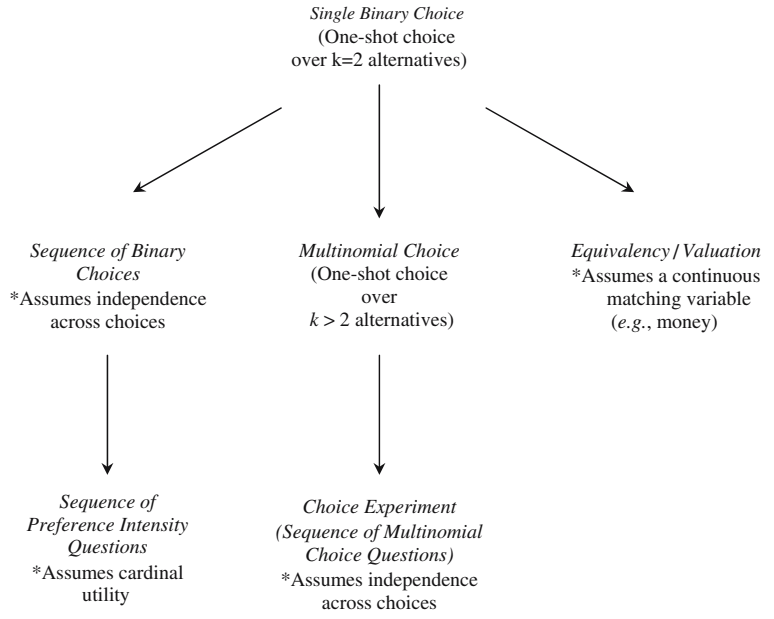

Fig. 1 Typology of elicitation formats

property for others. An example here would be a tax payment vehicle when the agent does not pay taxes.

3 A simple typology of elicitation formats

As seen in Fig. 1, preference surveys are typically undertaken in a number of standard formats.[13] The figure starts with a single binary discrete choice format. This is most basic format where the respondent is told about two different alternatives and then asked to indicate which is preferred.

The figure then notes the extensions of the single binary choice format. One extension is that of asking a *sequence* of binary choice questions. A number of formats can be shown to be strategically equivalent to this format, including the popular double-bounded dichotomous choice format in contingent valuation (Hanemann et al. 1991). A commonly used variant of this second extension is that of asking a sequence of intensity of preference questions which asks agents to rate one choice relative to the other on a numeric scale such as 1–10 (e.g., Johnson and Desvousges 1997).[14]

Another extension, asking a *multinomial* choice question, asks the agent to pick their most preferred out of $k > 2$ alternatives. A popular variant of this format is that of asking a sequence of multinomial choice questions (Louviere 1994) which is now often referred to the as the choice experiment format. A third extension, a *matching* question (sometimes

[13] Starred text in the figure denotes assumptions commonly associated with each format.

[14] We do not explore the properties of this type of preference elicitation question any further in this paper because of its need to make a cardinal assumption about utility rather than the traditional ordinal assumption of neoclassical economic theory.

called *equivalency* or *valuation* question), drops the attribute level (typically cost) of one alternative, and asks the agent to specify the quantity of that attribute level that would make them indifferent between the two choices.[15]

For each of these question formats it is possible to examine the divergence between the face-value response and the strategic response. It is also possible to look at differences in the type of information conveyed by different elicitation formats. Because it is the most basic, we start with an examination of the binary discrete choice format.

4 A single binary discrete choice question

A single binary discrete choice question, with one of the alternatives typically being the status quo, is one of the most commonly used preference elicitation formats and has a long history of use in survey research. Bishop and Heberlein (1979) showed that this format could be used along with a random assignment of different monetary costs to different respondents, to obtain the empirical distribution of willingness-to-pay or willingness-to-accept (WTA) values in a given population. Later papers by Hanemann (1984a, b) formally worked out the utility-theoretic approach from a random utility perspective (McFadden 1974); and Cameron (1988) provided a purely statistical approach to tracing out the latent (unobserved) WTP or WTA variable in a manner similar to dose–response experiments in biology or medicine. McConnell (1990), Kriström (1997), Haab and McConnell (1997, 1998, 2002) and Hanemann and Kanninen (1999) have also examined the statistical issues involved in using the binary discrete choice format. We do not address the substantive estimation issues raised in these papers, except to note that some of the implausible estimates that exist in the literature appear to stem from a failure to adequately model the data, or to incorporate sensible restrictions implied by economic theory.

Much of the attention focused on the binary discrete choice elicitation format in recent years is due to its recommendation by the NOAA Panel on Contingent Valuation (Arrow et al. 1993), due to its well-known property of being incentive compatible in many circumstances. Indeed, a core result in mechanism design theory, independently derived by Gibbard (1973) and Satterthwaite (1975), is that no response format that allows for *more* than two alternatives can be incentive compatible unless substantive additional restrictions are placed on agents' preferences.[16]

It has long been known that the binary discrete choice format is incentive compatible in some settings (Farquharson 1969). The best known examples are political races with only two candidates and binding approve/disapprove referenda. The NOAA Panel refers to the latter when making their recommendation to use a binary discrete choice format in contingent valuation (CV) surveys.

The first question to ask is whether it is the binding nature of a referendum that makes it incentive compatible. Carson et al. (1997) consider an *advisory* referendum,[17] in which the binding property is replaced with the more general property that the larger the percentage in

[15] The equivalency/valuation format is an extension of the single binary choice format in the sense that the agent is asked to provide the value of the missing attribute level that leads to a response of "indifferent" if the two alternatives were then provided as a single binary choice.

[16] This Gibbard–Satterthwaite theorem does not say that all (or even any) binary discrete choice formats *are* incentive compatible, only that all other formats are generically *not* incentive compatible.

[17] Many well-known referenda are technically advisory rather than binding. For example, Norway's vote on whether to join the European Union (EU) was an advisory referendum and some observers believed that if the referendum had passed by only a slim majority, the government would not have chosen to join the EU.

favor of the measure the more likely it is that the government will undertake the action. Even though they are not binding, such referenda are still incentive compatible.

The second question whether substituting an advisory *survey* for an advisory referendum alters the incentive properties of the mechanism? Green and Laffont (1978) have shown that any economic mechanism of the type being considered in this paper can be implemented using a statistical sample rather than a vote by the whole participation. Thus, we have:

> *Result:* It is possible to replace the binding nature of an incentive-compatible referendum with the more general assumption that the agency is more likely to undertake the action the higher percent in favor. It is also possible to replace a full public vote with a sample survey on the issue in question. Neither of these changes, alone or together, alters the original incentive structure of the binding referendum.

A small number of CV studies (e.g., Carson et al. 1987; Vossler et al. 2003), have in fact compared survey estimates to votes on actual binding referenda and have found them to be close. A large body of evidence also suggests that surveys taken close to the event generally provide good predictions of actual referendum votes.[18]

Two key assumptions have been made thus far in discussion. The first is that the agency can actually compel payment for a good if it decides to provide it. The second is that only a single issue is involved. Dropping the first assumption destroys the incentive properties of what we will call the *referendum/advisory referendum/advisory survey* (RARAS) mechanism. To see this, consider the case where a charitable organization wants to provide a public good through voluntary private contributions. A "yes" response to a binary survey question of the form: "Would you contribute $\$X$ to a fund to purchase this good if we started such a fund?" will encourage the charitable organization to undertake the fundraising effort. A rational agent would always want to say "yes" to such a question, even if their maximum willingness to pay was less than $\$X$, and even if they intended to pay less than $\$X$ (if anything) once the fund has been started.[19] In other words, this format actually gives agents an incentive to over-pledge, in order to obtain the opportunity to underpay. A number of empirical studies confirm this large divergence between survey-based predictions and actual contributions to such funds (e.g., Seip and Strand 1992; Champ et al. 1997; Foster et al. 1997).

Switching to the case of introducing a new private good does not improve the incentive situation. As long as there is any positive probability of wanting the new good at the stated price, the respondent should say "Yes — would purchase." The agent's logic is that such a response will encourage the company to produce the good, with the agent being able to decide later whether or not to purchase. Since increasing the agent's choice set in a desirable way increases utility, the optimal response is "yes." Folk wisdom from the marketing research literature supports the notion that consumers overstate their purchase proclivities for new products (Greenhalgh 1986). Evidence from experiments in economics (Cummings et al. 1995; Johannesson et al. 1998) also supports this conclusion. The marketing research approach has tended to either shift to a different measurement scale, such as the probability

[18] Predicting an election vote from a survey involves two difficulties unrelated to whether agents truthfully reveal their preferences in surveys. The first is that the information set used by voters on election day may have changed substantially since the time of the survey due to activities such as political advertising and media coverage, which is why surveys taken close to an election are generally more accurate than surveys taken earlier. The second is predicting who is going to actually vote. The characteristics of a random sample of the entire public may be quite different from the characteristics of the subset that actually votes.

[19] In many charitable fundraising efforts, the quantity of the good provided is increasing in the amount of money raised. In such a case, it may be optimal for an agent who desires the good to actually contribute a positive amount toward its provision (Bergstrom et al. 1986).

of purchasing (Inforsino 1986), or to ask about the preferred alternative from a set involving more than one good (Louviere 1994).

There is some irony in this result, as it has so often been maintained that if standard CV elicitation formats did not work well for private goods, they would work even worse for pure public goods, since the latter are not bought or sold in the marketplace and are thus less familiar to consumers. This argument is often used to justify prior experiments with private goods, to first learn how CV is likely to work in this more familiar (and hence "best case") scenario (Neil et al. 1994). But as seen above, the introduction of a new private good is one of the *worst case* scenarios for a binary discrete choice question. It should not be surprising that the binary discrete choice format, though it initially saw usage in marketing research, is now rarely used there.

The ability of the agency to coercively collect payment for the good is the property that causes the agent to try to influence the agency's decision in the desired direction taking account of both the cost and the benefits of the action to the agent.[20] Voluntary contributions allow for the possibility that the survey response encourages the fund-raising effort to be undertaken, and hence, the possibility of free riding during the actual fund-raising effort. Thus, agents, who want the good provided, should say "yes" (would contribute) to the survey question. In turn, it will be optimal for some of these agents to free ride in the expectation that other agents would contribute enough to provide the good. In this case, an initial survey "yes" response helps to set up the later opportunity to free ride with respect to the actual contribution.

For the private goods, a "yes" response (would purchase) to the survey encourages the production of the good while the agent gets to decide later whether to purchase the good. Thus, if the agent anticipates any positive probability of wanting to purchase the good, a "yes" response would be optimal. If the agent anticipates that the good will be offered irrespective of the nature of the responses but believes that the responses may influence the *price* of the good, then it is optimal for the agent to appear more price sensitive than they actually are. This result is often seen in marketing research where agents have been found to be more price elastic in surveys than in actual market purchases. What is interesting is that while the marketing literature has often noted the divergence between survey-based estimates and market behavior, it has failed to note the change in the sign of divergence with the change from a new good provision exercise to a pricing exercise. This may be because the emphasis in the new goods literature has focused on estimating purchase probabilities while the pricing literature has focused on estimation of price elasticities. The only problem with these two cases from the perspective of economic theory is not whether there should be a divergence between actual behavior and the survey estimate, but rather, whether the magnitude of the divergences empirically observed should be even larger.

[20] It is interesting to ask whether it is the two-step nature of a survey followed by a contribution/purchase that leads to the survey question not being incentive compatible. The answer is no. Consider the situation whereby the only way a public good can be provided is if it obtains the requisite plurality vote in a referendum and the legislature gets to decide whether to put the issue on the ballot for a vote. The legislature does not want to waste the public's time putting propositions to a vote if they stand little chance of passing. The legislature, or the measure's supporters, commission a survey to determine the likely fraction of the public that would vote in favor of it. The only consistent responses (given no change in the information set) to the survey and actual referendum vote are "yes" to both the survey and the referendum or "no" to both the survey and referendum. For those in favor of the measure, the only way to get the good is to get the referendum put on the ballot and have the measure passed. "Yes" responses to both opportunities increase the chance of both. For those opposed to the measure, saying "yes" to the survey increases the chance that it will get put on the ballot, which in turn increases the chance that the agent will have to pay for the good, even though the good is not worth the cost to the agent if provided.

There are other interesting implications of the lack of incentive compatibility of binary discrete choice survey questions for voluntary contributions and the introduction of new private goods with respect to other anomalies, such as insensitivity to the scope of the goods being valued. For instance, as long as the good is potentially desirable, it is optimal to say "yes" to the survey question. The scope of the good and its cost do not influence this decision unless the good becomes so small that even at a zero cost it is not desired, or the cost becomes so high that it would never be purchased. In both of these latter instances, either a "yes" or a "no" response by the agent will have the same effect on their utility.

If the binary choice is between two different forms of a quasi-public or private good, then desirable incentive properties can be restored as long as only potential users are interviewed.[21] To see this, consider the example of a campsite which is currently unimproved and currently has a low (possibly $0) entrance fee. The alternative is to improve the campsite and increase the entrance fee. The agent should now choose between the status quo price/quality combination and the alternative price/quantity combination in order to maximize utility. This binary choice can be shown to have identical properties to the RARAS survey mechanism. The property that this mechanism needs in order to be incentive compatible is the ability of the agency to subsequently force one of the alternatives on a particular agent, irrespective of that agent's preferences. Two important caveats should be kept in mind. First, in this situation the total number of times the good will be used under the alternative is endogenous. In our campsite example, if the higher quality-price campsite alternative provides more utility than the status quo, the anticipated number of visits to that campsite under that alternative may be larger or smaller than under the status quo. Second, for agents whose probability of using either configuration of the good is zero, any response has the same impact on their utility. This problem is not usually seen, since most recreational surveys are either done on site or from lists of users. Similarly, marketing researchers typically screen out non-users of a product class before asking preference questions.[22] The risk in both instances is that focusing on current users of the good will miss those who would likely use the good if its quality were improved or its price reduced.

This choice between two alternative configurations of a good works for both public and private goods, irrespective of the nature of the payment obligation, as long as the agent desires the good at no cost. To see this, consider a private charity that wants to build one of two different monuments in the center of town. The charity conducts a survey to determine which monument is preferred. The higher the level of support for a particular monument, the more likely that monument will be built. The agent should pick their preferred monument, since this increases their utility more than the alternative monument and neither monument imposes any cost on the agent. Our favorite example of a private good question is the bar owner who surveys patrons and asks whether they would prefer to have the bar's sole draft beer, currently a domestic brand priced at $1, switched to an imported brand at $2. The bar patron should pick the import only if having that alternative available provides more utility

[21] Quasi-public goods are those provided by the government but for which it is possible to exclude members of the public from its use. This exclusion can occur in terms of charging a price to use the resource, having the agent spend money or time to use the resource, or by having the resource bundled as an attribute of a privately purchased product. Common examples include government campgrounds and houses located near public lakes.

[22] There are exceptions. For instance, Boxall et al. (1996) ask hunters in Alberta about two different management/ cost regimes, for an area that few currently hunted in or were likely to hunt in under the alternative scheme. The contingent valuation estimate was dramatically larger than the travel cost estimate, which is fairly unusual in such comparisons for quasi-public goods (Carson et al. 1996). When the estimate of the change in the probability of use is used to adjust the CV estimate, however, the two approaches result in quite similar estimates.

than the domestic. Note that the number of beers that will be purchased is not revealed by the agent's choice and could go up or down.

The second key assumption in the discussion of the RARAS mechanism is of a single up/down vote on a single issue. It is similarly not possible to relax this take-it-or-leave-it condition, and there are several common instances where it is violated. The best-known ones are the rules for school bond referenda in many areas (Romer and Rosenthal 1978; Lankford 1985). The school board gets to propose the level of educational inputs and the tax rate. However, if the referendum is voted down, the school board can only bring up another referendum measure with a level of educational inputs and a tax rate that is lower than those voted down but higher than the default status quo. A respondent who prefers the initially offered bundle to the status quo may nonetheless have an incentive to vote against it in order to gain opportunity to vote in favor of an even more preferred provision/tax package. With respect to valuation of an environmental project, Richer (1995) showed that his CV WTP estimates were influenced by information about whether a different alternative plan for a national park in California's Mojave Desert was likely to be put forth if the current plan described in the survey was not approved. Another variant is where there is another party (e.g., another government agency or private entity) who potentially can provide the good.[23] The general principle is that direct linkage between a decision on one issue and a decision on another issue can cause difficulty in interpreting the result, as the optimal response of the agent should generally take the sequence of decisions and options into account.

Table 1 summarizes the incentive properties of binary discrete choice questions by the type of good and the payment characteristics under the assumption that agents that the take-it-or-leave it condition holds.[24] What is striking is that anomalies with respect to a divergence between estimates based on stated preferences and estimates based on behavior are heavily concentrated in the two cases that are not incentive compatible.

There is a further condition that is important for the interpretation of the results but not for the incentive properties of the RARAS mechanism. The agent needs to believe that if the agency implements a particular alternative: the specified good Q will be provided and the

[23] This problem appears to have influenced the Cummings et al. (1997) results. In that experiment, agents are randomly assigned to a "hypothetical" treatment and a "real" treatment in which they vote on whether to contribute a specified amount per agent to provide the good. The estimate based upon the hypothetical treatment is higher than from the real treatment, though Haab et al. (1999) argue that the significance of the difference depends upon how the larger variance in the "hypothetical" treatment is taken into account. We believe that to many of the agents they interviewed in Georgia, the Cummings et al. hypothetical treatment should have appeared as an attempt to determine if it was possible to mount a fundraising effort to provide printed information booklets on toxic hazards to poor people in New Mexico. Thus, we would have expected the hypothetical treatment WTP to be higher than true WTP. However, uncertainty about why Georgians should be asked about contributions to a New Mexico program may have led to the larger variance found by Haab et al. (1999). For the "real" treatment we would have expected an underestimate of true WTP due to the possibility of having some other group pay to distribute the booklets. A later experiment by Cummings and Taylor (1998) effectively replicates this experiment but with additional treatments where there are different probabilities that the vote taken by the group is binding. The WTP estimate decreases progressively from the "hypothetical" treatment to the "real" treatment as the probability that the group vote is binding goes from 0 to 1. This is the result our model predicts if all treatments were perceived by agents as being consequential and there are competing incentives to over-pledge and free ride in all of the probabilistic treatments. The key prediction, if there was no incentive to over-pledge in the "hypothetical" treatment and free ride in the "real" treatment, would be that all treatments with a positive probability of the vote being binding should result in similar WTP estimates. Carson et al. (2004) explore this issue at length and show that the theoretical prediction holds empirically in a carefully designed experiment with a public good.

[24] Note that the while incentive properties here do not depend upon whether all agents find the good not to be a bad at zero cost most econometric analysis makes this assumption in order to recover the WTP distribution. This property may hold in some instances like development versus preservation issues where there are usually both gainers and losers from the proposed policy as well as those who are indifferent at a zero cost.

Table 1 Incentive properties of binary discrete choice questions

Type of good	Incentive property
New public good with coercive payment	Incentive compatible
New public good with voluntary payment	Not incentive compatible
Introduction of new private or quasi-public good	Not incentive compatible
Choice of which of two new public goods to provide	Incentive compatible
Change in an existing private or quasi-public good	Incentive compatible, but choice does not reveal information about quantities

stated price P will be assessed. If instead the agent believes that Q* will be provided and price P* assessed should the alternative be chosen by the agency, then their optimal response should be based upon (Q*, P*) rather than the stated (Q, P). Note that this condition holds for interpreting actual votes or actual consumer purchases as well as for responses to preference survey questions.[25] An important implication of this condition is that if the goods and prices used in a preference survey go beyond what the agent finds plausible, then the preference survey question is likely to be answered on the basis of the expected good and the expected price rather than the stated ones.

4.1 Implications of cost uncertainty

Binary discrete choice preference surveys often provide a cost (in monetary or other terms) for each alternative and this cost information plays a key role in estimating welfare measures. What role should an agent's uncertainty over cost play in the answers they give? The answer is obvious if, say, a survey provides a cost estimate of $\$X$ but the agent thinks that since the government has a proclivity for cost overruns the actual cost will be double that value. The analysis should be performed with the cost as perceived by the agent.

A more interesting case is when the agent takes the survey and treats $\$X$ as the expected value of some probability distribution around $\$X$. Here the key issues are whether the original status quo choice set will still be available and whether a commitment to pay for the good is required before the cost uncertainty is resolved. These two conditions determine whether shifts from an original "yes" to a "no" or vice versa are possible given a mean-preserving increase in cost uncertainty. Table 2 displays the possible outcomes.

Consider first the case of the provision of some public good with a coercive payment mechanism where the status quo choice set will still be available but where one has to commit *ex ante* to paying the uncertain cost. This commitment translates into income uncertainty, and hence is never preferred by risk adverse agents. Therefore one would expect to see some shifts from "yes" to "no" responses but no shifts in the opposite direction. As such, location statistics of the WTP distribution like the mean and median will shift upward relative to the case with no cost uncertainty. The second case, case where an *ex ante* commitment is required but the status quo will no longer be available, leads to a similar result.

[25] Carson et al. (1994) show, for instance, in a recent CV study in California that respondents who do not currently pay taxes are willing to pay more than respondents with otherwise identical characteristics. Respondents who believe that the state government would assess the "one time" tax in multiple years are willing to pay less than respondents who think the fee will only be assessed once and respondents who don't believe the plan will work completely are willing to pay less than those who think it will work. See Randall (1994) for a discussion of this issue in the context of the travel cost model. There are large literatures in marketing and political science dealing with what are effectively the P's and Q's perceived by agents when they make decisions.

Table 2 Effect of increased cost uncertainty upon binary choice

	ex ante choice (i.e., commitment)	*ex post* choice (i.e., no commitment)
Status quo still available	Can only shift Yes → No	Can only shift No → Yes
Status quo no long available	Can only shift Yes → No	Can shift either Yes → No or No → Yes

In the third case, when the choice can be made *ex post* after the uncertain cost is observed and the status quo choice set will still be available, increased cost uncertainty can lead to possible shifts from an original "no" to a "yes" response. The main examples of this are provision of a pubic good via voluntary contributions and the introduction of a new private good. The basic logic in this case is that since the status quo choice set will still be available, agents will either favor or be indifferent to the addition of the new alternative. Increasing the level of uncertainty can cause some agents who had been indifferent to adding an alternative to now favor it. Changes from a "yes" to a "no" response cannot occur, although it is possible that an increase in cost uncertainty can make some agents who are already in favor to be worse off.

The last case occurs where only *ex post* commitment is required and the original status quo choice set will no longer exist if the alternative is provided. The main examples here are quasi-public goods and private goods where only one of two possible configurations of the good will be offered (e.g., a low quality/low price recreation site versus a high quality/high price version of the site). In this case, it is possible that increasing the degree of cost uncertainty results in both shifts from "yes" to "no" as well as from "no" to "yes."

There are a number of other informational issues we do not explore, except to note that a formal analysis of the different types of uncertainty is likely to be more productive than the frequently invoked vague concept of agent unfamiliarity with a good as a justification for all types of apparent aberrant behavior. Much of the richness of economic theory in recent years has come from the introduction of different types of uncertainty and asking how agents should optimize in the face of it (Varian 1992). Particularly relevant is uncertainty over the probability of provision and its interaction between the scope of the good and the method by which it would be provided, which can easily produce the appearance of respondents being willing to pay more for less. There is also a growing literature on how agents process information in elections and referenda (e.g., Popkin 1991; Lupia 1994). This literature suggests ways in which agents make reasonably informed decisions based on imperfect information. Further, simply providing more information does not necessarily lead agents to make decisions closer to those they would make if fully informed (Lohmann 1994).[26] This suggests that the informational content of a survey used for environmental valuation should be examined to see if agents were given a reasonably complete, comprehensible, and balanced presentation of the alternatives offered.

5 Generalizing the binary choice question format

There have been a number of attempts to generalize the binary discrete choice. These have usually had two purposes, tightening the confidence interval on the statistics of the WTP distribution

[26] Consider an agent who initially favors a project, believing both its benefits and costs to be small. The agent would still favor the project upon learning the reality, namely that both its benefits and its costs are large. If informed, however, that the cost is large but not given the corresponding benefit information, the agent will oppose the project. Much advertising in marketing and political campaigns operates on this notion of providing selective "half-truths."

for a single good or providing information about the WTP distributions for a variety of related but different goods in order to help decision makers pick the best option. The most popular approach used to tighten confidence intervals is the double-bounded dichotomous choice format (Carson 1985; Hanemann et al. 1991). The most popular formats for providing information about multiple goods are the sequence of pair comparisons (Magat et al. 1988; Peterson and Brown 1998) and choice experiment formats (Adamowicz et al. 1994; Louviere et al. 2000).

There are two key features of any approach that asks for a discrete response to other than a single binary discrete choice question that influences its properties. The first is that the information derived can always be decomposed into a response to two or more binary discrete choice questions. This is obvious for double-bounded questions and the sequence of paired comparisons as these formats simply consist of asking multiple binary discrete choice questions.[27] For the choice experiment, this can be seen by looking at a single multinomial choice question with three alternatives (A, B, and C) and seeing that a respondent who has indicated their most preferred option (e.g., A) has really revealed that they prefer A to B and A to C. No information is provided about whether B to C without explicitly asking for a choice to be made between those two alternatives. Extension to more than three alternatives or asking for preferences in multiple choice sets does not change the fact that the preference information obtained is really that about different binary choices. It is well known that offering respondents a choice between more than two alternatives violates the Gibbard-Satterthwaite necessary condition for incentive compatibility and it is typically easy to work out cases where it is not in the interests of some agents to truthfully reveal their preferences. The second is that asking the agent explicitly or implicitly about more than one pair of alternatives provides an additional piece of information that can change how the agent answers relative to a single binary choice question.

We take up double-bounded dichotomous choice questions, sequences of paired comparisons and multinomial choice experiments in turn looking at the likely incentive and information effects that may be present.

5.1 Double-bounded dichotomous choice questions

The inherent problem with a binary discrete choice question is the limited information that a response provides about the agent's preferences.[28] Double-bounded dichotomous choice estimators have become popular in the environmental valuation literature because they tend to dramatically shrink the confidence intervals around point estimates of parameters of the willingness-to-pay distribution. The approach is straightforward: If an agent says "yes" to the initial cost amount asked, then ask them the same question at a higher amount, and if the agent says "no" to the initial amount, ask the same question at a lower amount.[29] Initial versions

[27] For a sequence of pair comparisons, the sequence may be of any number of paired comparisons beyond the initial pair. It is also possible to expand the double-bounded concept to asking three or more questions at different cost amounts (e.g., Bateman et al. 1995).

[28] The only information provided is whether the agent's WTP for the good is higher or lower than the single amount asked about in the survey question. It is possible to use parametric assumptions about the underlying WTP distribution to effectively overcome this sparse information but such assumptions can play a large role in the estimates derived. Non-parametric approaches to the use of binary discrete choice data (e.g., Kriström 1990) exist that make the power of these assumptions abundantly clear.

[29] The double-bounded model bears some resemblance to the iterative bidding game approach used in the early CV literature (Randall et al. 1974), which was often found to suffer from the phenomena of "starting point bias", in which the amount initially stated influences the agent's final WTP amount. There are some key differences, however, which make the two approaches fundamentally different. The initial cost in the iterative bidding game was never intended to reveal information about the good's actual cost and the iterative steps from that amount

of the double-bounded format relied on double sampling/interval censoring statistical models (Carson 1985; Carson and Steinberg 1990; Hanemann et al. 1991). They assumed that agents have a single latent WTP value and that the responses to both the first and the second questions are based upon comparing this latent WTP value to the stated cost amount in each question. Statistically, the implication of this assumption is that, with appropriate conditioning, there is perfect correlation between the WTP distributions implied by the responses to the two questions.

Following Cameron and Quiggin's (1994) pioneering examination of this assumption, several stylized facts have emerged concerning the comparison of WTP estimates based upon the first binary discrete choice question versus the second question: (a) WTP distributions implied by the first and second questions are not perfectly correlated, (b) the WTP estimate based upon just the first question is higher than the WTP estimate based upon both questions, and (c) the number of negative responses to the second question is higher than would be expected based upon the WTP distribution estimate from the first question alone. Herriges and Shogren (1996) have put forth a model for the second question based on starting point bias, Alberini et al. (1997) have put forth a general error-components model, and McLeod and Bergland (1999) have put forth a Bayesian preference-updating model to handle these issues.

What sort of effects should the asking of a second binary discrete choice question have on the latent WTP distribution?[30] From our perspective, the key property of this format is the prediction of the empirically observed stylized fact (a) that the responses to the two questions are not perfectly correlated. Any interpretation of the information signal provided by offering to make the same Q available at two different prices implies that less than a perfect correlation between the two responses should be observed. Beyond this prediction it is necessary to make more specific assumptions about agent beliefs.

The best-case scenario here is that the agent takes the second price as the expected price but now considers the price to have some uncertainty surrounding it.[31] Consistent with the discussion in the previous section, parameters such as mean or median WTP will be shifted downward in the second question for risk adverse agents and public goods, even though preferences for it have not changed. This (as well as other belief structures) will produce the second stylized fact that the standard location statistics of the WTP distribution implied by the second question are shifted downward relative to those implied by the response to the first question.

There are, however, several other plausible alternatives for what asking the second price might signal to agents. One of these is that the agency is willing, in some sense, to bargain over the price. For agents who originally answered "no" and were asked a lower price, the optimal response may be to answer "no" again in hopes of being offered an even lower price. This should result in the second WTP response being "no" for some of these agents, even

Footnote 29 Continued
were usually quite small. In contrast, the statistical tools used to analyze data from both the binary discrete choice and the double-bounded discrete choice formats exploit the agent's conditioning on the cost numbers explicitly provided and the difference between the first and second price is usually much larger. Most good studies using a double-bounded format go to some effort to provide a rationale to the agent as to why the cost number used in the second question is different from that of the first. An interesting variation on the double-bounded format is a single binary discrete choice format with a follow up open-ended question. Farmer and Randall (1996) analyze this format from a theoretical and empirical perspective and obtain results similar to those described here for the double-bounded estimator: the second responses tend to be biased downward.

30 It is possible to design experiments to investigate the relative importance of particular effects in particular contexts (e.g., Burton et al. 2003).

31 Alternatively, if the agent thought the first price had some uncertainty surrounding it, asking the second price might increase this level of uncertainty, since for the double-bounded estimator the first and second prices are typically fairly far apart.

though had the amount been asked at the first question the response would have been "yes." A similar effect can be found with respect to those who originally answered "yes." Since the good was originally offered at a lower price, it can presumably be provided with some positive probability at the initial price. As such, some agents will find it in their self-interest to risk not getting the good by holding to the lower price and saying "no" to the second higher price, even though their WTP exceeds the second price. The effect of such behavior would be to lower the WTP distribution implied by the second question, and hence, reduce estimates of the mean and median WTP.

Another plausible assumption is that the actual cost believed by the agent will be some type of weighted average of the two prices. If this assumption is made, the second question should be answered on the basis of this weighted average. It is straightforward to see that for an initial "no" response, any weighted average of the first and second prices is higher than the second price. For an initial "yes" response, any weighted average of the first and second prices will be lower than the second price.[32]

The last plausible assumption we consider is that the agent might interpret the signal given by the second price as implying that the quantity has changed to match the changed price. For an initial "no" response, the shift in quantity that is consistent with the reduction in price is to reduce the perceived quantity/quality of the good that would be provided. The implication of this is to shift the WTP distribution implied by the second response to the right for these respondents. This is a commonly voiced concern in focus groups and debriefing questions. For agents who initially said "yes", the shift in perceived quantity is upward. There does not appear to be any collaborating evidence to support the proposition that this is a common phenomenon.

What should be grasped from this discussion is that, to a rational agent, the appearance of a second price must signal that something is going on. All of the plausible assumptions lead to the key prediction that the correlation between the WTP distributions implied by the two questions is less than unity. More specific predictions require more specific assumptions about beliefs. Most plausible assumptions also shift the WTP distribution implied by the second question to the left for agents who initially gave a "no" response, and hence, produce an "excess" number of no/no responses. For agents initially giving a "yes" response, it is possible for the WTP distribution implied by the second question to shift either to the left or the right, but only the price averaging assumption has much credence in terms of the possibility of producing an upward shift in the WTP distribution.[33] On balance, we would expect the WTP estimates from a double-bounded format to be smaller than those from a single-bounded format. This hypothesis tends to be strongly supported by the empirical evidence. It may still be desirable to use the double-bounded format in CV studies; however, this desirability rests on the analyst's tradeoff between the likely downward bias and a tighter confidence interval (Alberini 1995).

[32] Note that this assumption is not inconsistent with the arguments concerning uncertainty and the two may be combined. For initial "no" responses, this effect of adding uncertainty is reinforcing in a downward direction. For initial "yes" responses, the effect is in the opposite direction and mitigates the upward effect of price averaging.

[33] It is possible to get non-truthful yes-yes responses in an instance where the payment vehicle is not seen as binding and the respondent desires the good. The classic example is asking for a voluntary payment but it can also occur when, for example, in the case of asking about an increase in a particular tax that a particular agent doesn't have to pay. What is important to note here is that the incentive structure has not changed between the single and double-bounded choice format in this instance, although the fraction of agents in the sample for whom the payment vehicle is not binding may be more identifiable with the response to the second question, particularly since the response from those agents for whom the payment mechanism is binding will tend to be biased downward.

5.2 Sequences of paired comparisons

To value multiple goods it is necessary to ask agents to make tradeoffs concerning them. The single binary discrete choice question effectively asks an agent about a status quo good with a price (which is often implicitly zero cost increase) and an alternative good at a different price (typically, higher if the alternative good is perceived to be better in some sense). The simplest version of a sequence of paired comparisons asks multiple binary choice questions of different goods, often related, against the same status quo. The main additional problem which creates incentives for some agents not to truthfully reveal their preferences arises from their beliefs about how the agency will treat, that is aggregate, the information across the different paired comparisons.

In an ideal world in which the objective involves valuing different public goods, the analyst would like the agent to treat each paired comparison independently. If this happened, the desirable properties of a single binary discrete choice question with a coercive payment requirement can be repeatedly invoked. There is a very simple question, however, which illustrates the fundamental difficulty with a sequence of paired comparisons. Consider the case of air pollution levels in a city. The agent is asked to rank different pairs of air pollution levels, which involve different costs and different health effects and visibility levels. Since air pollution in the city is a public good, however, all agents must face the *same* air pollution level. If $k$different air pollution levels are described to the agent in the course of the sequence of paired comparisons, the agency must have some method of choosing among the k different levels. *Any* particular method that the agent perceives that the agency is using to incorporate agent preferences into its choice of an air pollution level generally will provide an incentive for some agents to engage in non-truthful preference revelation.[34] In some instances, it will even be optimal for agents to reject their most preferred level (out of the k) in a particular paired comparison. Once this is possible, standard methods of inferring value from choices no longer work. The essential problem is that an agent's optimal choice depends both upon their own preferences, expectations about what other agents will do, and the perceived rule for aggregating the results of each paired comparison. This result has long been established in the literature on the properties of voting rules (Moulin 1994).

With quasi-public and private goods, the difficulties noted for public goods still exist, with the exception that it may be possible for more than one of the k goods to be provided. This possibility tends to reduce the likelihood that an agent will make a choice that is not their favorite, and in the following section, we discuss the aggregation issue in this situation further.[35]

[34] That is any method short of a credible pledge to throw away all of the information provided in the sequence of paired comparisons except for one randomly chosen pair. However, such a credible pledge cannot be made in the context of a survey. This issue is discussed further in Sect. 6 below.

[35] There are further issues related to a sequence of paired comparisons that need to be addressed in any particular analysis. The first of these is the strong possibility that the scale term associated with each paired comparison is different. If this is the case, then much of the gain in precision and the ability to deal with changes in attributes associated with asking the sequence of paired comparisons may be much less than expected. The second is that most rules for combining information from different paired comparisons implicitly require that the Irrelevance of Independent Alternatives (IIA) assumption to hold. This property is routinely rejected in paired-comparison data. The third involves the common use of pairs where both alternatives are off of the agent's current utility frontier and neither represents the status quo. This practice requires much stronger assumptions about the nature of the agent's utility function than is typically assumed in order to combine the data from different paired comparisons.

5.3 Multinomial choice format and choice experiments

The term "choice experiment" is sometimes used to refer to a sequence of paired comparisons but it is more typically used to refer to a sequence of multinomial choice questions. All of the issues discussed previously about a sequence of paired comparisons carry over to the sequence aspect of a sequence of multinomial choice questions. As such, we first concentrate on the issue of the strategic issues that an agent faces when answering a multinomial choice question, that is picking the most preferred alternative out of $k > 2$ alternatives. The fundamental issue is how agent perceives the agency translates the responses to the multinomial question into actions. The simplest case consists of generalizing the decision rule used in the binary discrete choice format by assuming that the agency will provide only one of the k goods, and the higher the percentage of the sample picking any particular alternative, the more likely that alternative will be provided. The well known result from the voting literature on multi-candidate races with a simple plurality winner is that an agent's strategy reduces to a binary choice between the two alternatives that they believe will receive the most votes, independent of the agent's actual first choice. The rationale behind this result is straightforward: only the top two alternatives have a chance of winning; picking the most preferred alternative among these two will maximize the utility of the agent's final outcome.[36] The agent is truthfully revealing their preferences, but such truthful preference revelation is, as it should be, conditional on expectations over the choices of the other agents. However, the agent may not be answering the question of interest to the analyst, since it will be optimal in some instances for the agent to pick an alternative other than the (unconditionally) most preferred one.

Let us now consider perhaps the opposite case, one of particular relevance to private and quasi-public goods, by changing one of the key assumptions. Now, instead of only one of the k goods being supplied let $k - 1$ of the goods be supplied. To keep matters simple, assume further that the agent uses at most one of these goods. Examples of such a choice context might be a computer company that was going to offer four out of five configurations of a particular computer model, or a government agency that has to close one out of five recreational fishing lakes, leaving the remaining four open.[37] In this case, it is optimal for the agent to pick the most preferred alternative out of those offered. Formally, it can be shown that this case collapses to a binary discrete choice of the agent's most preferred alternative against another stochastically chosen alternative. To see this, note that the worst possible outcome for the respondent is that the agent's first choice is not made available. But because all the other alternatives would be provided, the agent's second choice would be available. Effectively, this is a determination of what alternative will not be provided. In pairing the agent's favorite alternative against *any* of the other alternatives, the agent's optimal response is to pick their most preferred.

[36] With a richer model of agent expectations, it may be optimal for the agent to vote for an alternative that is not one of the top two if there is enough uncertainty over the expected finish of alternatives and the utility differences between the alternatives is large enough. The manifestation of this proposition can be seen in the behavior of fringe political candidates in plurality winner elections. Such candidates try to convince voters that they have a non-trivial chance of winning, that the difference in positions between the two front-runners is extremely small, and that they are much closer to the voter's ideal point.

[37] Other permutations may be useful in practice. For instance, if there are four current recreational sites, a fifth new one with a specific set of attributes, could be added to the choice set and the agent told that only four of the five would be provided with one of the existing sites closed if the new site was opened. This avoids the problem that any addition to a choice set that has a positive probability of being used in the future is desirable as long as the current options are also going to be available.

The general result is that if all but j of the alternatives are to be provided, then the alternative chosen by the agent should be one of their j favorites. Often the number of alternatives that will be provided is unknown to the agent at the time of making the multinomial choice. A stochastic version of this result has the agent trading off the utility of sets of alternatives with different maximum elements against their own prior on j, and their prior on the choices made by other agents. Doing so reveals that agents will pick either their (unconditionally without considering the responses of other agents) favorite alternative or close to it, as long as one of three conditions holds: the expectation of j is fairly small, the utility difference between the agent's most favorite alternatives and the other alternatives is large, or the prior on the choices by the other agents is fairly uninformative. The implication of this is that agents will appear to make mistakes or optimization errors more often. If they don't pick their favorite, they should pick an alternative close to it.

The statistical manifestation of this type of behavior is a violation of the error term properties associated with the Irrelevance of Independent Alternatives (**IIA**) assumption. In empirical applications of this elicitation format, the IIA assumption is usually violated. While there are a number of other good reasons for this assumption being violated, such as the rationale behind the classic red bus/blue bus problem (McFadden 1980), it is typically impossible to separately identify the reason for an IIA violation.

To uniquely recover the latent WTP distribution, it is necessary to have an estimate of the correct scale factor.[38] Unfortunately, the optimal strategic behavior in this case is often observationally equivalent to direct manipulation of the scale parameter, making recovery of the correct scale factor impossible.[39] While it is at times asserted that a choice experiment is somehow too complex for an agent to strategically misrepresent their preferences, this is clearly not the case. The agent's optimal strategy is usually simply to try to induce the agency to supply the good with the most desired set of attributes at the lowest price. This can be done by picking something relatively close in attribute/utility space when the price of their favorite alternative seems "too high" as judged by the pattern of prices observed previously which suggests that the preferred alternative could be delivered at a lower cost.[40] The agent wants to appear to have an infinite demand elasticity at this cost and to be uninterested above it.

Tests for whether data from stated preference surveys and revealed preference observations are consistent with each other and can be combined after (potentially) allowing for a difference in the scale factor (Adamowicz et al. 1994; Swait and Louviere 1993) are commonly passed and have sometimes been interpreted as ruling out non-truthful preference revelation. However, such tests are really tests against random responses in the stated preference data, not tests against intentional non-truthful preference revelation. These tests are still important though because they help to rule out the extreme preference lability that is sometimes asserted by psychologists. Indeed, if this phenomenon was generally present one would not expect to see the consistent acceptance of the "scalability" of revealed and stated preference data (e.g., Louviere et al. 2000).

Unfortunately, to obtain an unbiased estimate of total value it is necessary to have a reasonable estimate of the correct scale factor. The situation may be different though when looking at the marginal tradeoffs between attributes (e.g., Carlsson and Martinsson 2001).

38 The scale parameter is typically taken to be the negative inverse of the price coefficient.

39 For quasi-public or private goods, the correct scale factor can sometimes be obtained from a model estimated using only the choices currently available in the market. This will not typically be the case for public goods.

40 Note that such behavior can clearly result in the appearance of initial "learning" as one goes through the choice sets, with most of the learning occurring in early choice sets. The difference is that it is learning about costs and strategies rather than learning about preferences.

Here if agents follow simple strategies focused on the cost variable such as the one noted above, it will still be possible to obtain useful estimates involving marginal tradeoffs between attributes.[41] That is because the scale factor cancels out when making marginal comparisons. The robustness of marginal comparisons to a variety of misspecification issues has long been noted (e.g., Ruud 1983) and it is not surprising that studies that find close correspondence of behavior suggested by a choice experiment and actual behavior typically look at marginal comparisons.

With either subadditivity or superadditivity of the utility, and $k - j\,(j > 1)$ alternatives to be provided, it is possible to find conditions where it is optimal for some agents to indicate their (unconditional) least preferred alternative. The rationale here is that the agent's outcome utility is defined on the set of goods to be provided, not the individual goods taken independently. This is a hopeless situation for learning anything reliable about agent preferences for individual goods.[42]

An alternative to asking agents to pick their single most preferred alternative out of k is to ask them to rank-order all k alternatives. This exercise could potentially provide considerably more information, but an analysis of the agent's strategic incentives becomes considerably more difficult. The same issue for the agent still exists: how does the agency translate the ranks into a choice of which of the k alternatives to provide. Methods for dealing with rank data in a manner consistent with economic theory effectively require the IIA assumption to hold for all possible subsets of the ranked data. This implies that it is possible to explode the data to form sets of multinomial choice questions down to all of the implied binary comparisons (Chapman and Staelin 1982). The IIA assumption can be tested but it does not appear to generally hold for contingent ranking data and welfare estimates can be substantially impacted if the IIA assumption does not hold (e.g., Hausman and Ruud 1987).[43]

6 Equivalency/valuation question formats

Ideally one would like to have the agent's actual WTP or WTA rather than a discrete indicator of it. So it is not surprising that many early CV studies used an open-ended direct question.[44] Many economists thought that these early efforts would fail because agents would give extremely high WTP answers. This did not happen (e.g., Brookshire et al. 1976), and interest in survey-based valuation methods grew in part due to this finding.

[41] Other forms of strategic behavior are, of course, possible but are often less obvious than those suggested by offering the same or quite similar goods at very different prices in the same or different choice sets.

[42] The simple dinner options example serves to illustrate this. Consider three alternatives, a hamburger, fish, and chips where the agent gets to indicate their single most preferred alternative and thinks that two of the alternatives are likely to be provided. Even though chips are the least preferred alternative, if only one alternative is to be provided, it may well be optimal for an agent to indicate chips as their first choice if they prefer both a hamburger with chips and fish with chips to the possibility of a hamburger with fish.

[43] A major problem occurs when there are a group of respondents who do not appear to want to trade off one of the attributes against money. The appearance of such lexicographic preferences (which may be an optimal strategy under some conditions) can lead to infinite WTP estimates. A subtler problem occurs in that the variance of the error term appears to be substantially larger for "middle" ranks than the most and least preferred alternatives. The need to model the differences in the scale terms across the ranks reduces the gains from asking for a complete ranking. This suggests a strategy that asks an agent for their "best" and "worst" alternatives may extract the much of the useful information while minimizing the agent's effort.

[44] The continuous-response format is known as a "matching question" in the psychology literature, and is a special type of open-ended question in the survey research literature.

The early problem that researchers *did* find with the direct question was that agents often wanted to know what the project would cost them. Agents did not understand why they were not provided the cost information if the agency had worked out the details of how the good would be provided. Further, some agents appeared to have great difficulty formulating a (continuous) WTP response. This led to very high non-response rates and a large number of so-called "protest zeros" which were typically dropped from the analysis; and further, to speculation that survey respondents did not have "well-defined" preferences in an economic sense.

Three different directions have been tried to overcome this problem. The binary discrete choice format discussed above avoids one of the key problems by giving agents a specific cost number and then uses a statistical analysis that "appropriately" conditions on agents' reaction (favor/not favor) to that value (Bishop and Heberlein 1979). The earlier iterative bidding game method suggested an initial amount, then iterated up or down from that amount in small increments (Randall et al. 1974). The so-called "payment card" approach asks agents to pick a number (or any number in between) from an explicit list of values (Mitchell and Carson 1986, 1989; Cameron and Huppert 1991). The latter two methods can come close to achieving a WTP response in continuous terms; and, except when these formats have special properties, the discussion of the continuous-response format will apply to these formats as well.

With different elicitation formats came the inevitable urge to compare their results (e.g., Smith and Desvousges 1986). Researchers were dismayed to find that different response formats led to different WTP estimates, and the divergence between these estimates is frequently cited as one of the major reasons why estimates based on stated preference questions should be rejected (Hausman 1993; McFadden 1994).[45] The stylized fact here is that discrete choice formats produce higher WTP estimates than do continuous-response formats (e.g., Boyle et al. 1996).

Should the divergence in estimates from different formats be surprising?[46] No, given the Gibbard-Satterthwaite result, it is impossible to formulate a continuous-response question that has the same incentive and informational properties as an incentive-compatible binary discrete-choice question. Many researchers looking at the results, however, have been misled by the face-value dilemma. The divergence between the estimates from the different formats suggested that either agents were not truthfully revealing their preferences to one or more of the elicitation formats not that they did not have well-defined preferences in the sense used by economists.

As noted above, the expectation of many economists was that most agents would provide very large WTP responses when asked an open-ended WTP question if they were acting strategically but not truthfully.[47] However, the opposite phenomenon was observed: estimates from binary discrete-choice questions were higher than those from continuous-response CV questions and whose continuous-response CV questions contained a high fraction of zero responses.

[45] The irony in this position is that estimates of other economic quantities based upon substantially different econometric techniques have typically differed even though data on actual behavior was being used. The usually recommended approach in this situation has not been to discard economic theory and econometric methods, but rather, to understand the source of the differences.

[46] According to some cognitive psychologists, the divergence between framing provided by binary choice and open-ended matching question is a key problem with economic theory (Tversky et al. 1990).

[47] It is interesting to note here though that Samuelson (1954) correctly foresaw survey respondents revealing an amount lower than their true WTP.

Faced with an open-ended question, a very large WTP response does turn out to be the optimal strategy for an agent who believes (a) the cost of the public good to the agent is fixed, (b) their true willingness to pay for the good is larger than the expected cost if provided, and (c) the good is more likely to be supplied the larger the sum of the willingness-to-pay responses given by agents. Condition (a) requires the agent to believe that the agency will not extract any of the consumer's surplus either for itself or to shift the burden to paying for the goods to those with higher willingness to pay for it. Condition (c) corresponds to the benefit-cost criterion but it is hard to find a single instance where an agency decision has been made based purely on that criterion. There is little evidence to suggest that agents believe that the agency is simply summing their WTP responses. As such, we believe it useful to consider a variety of other beliefs that agents may hold.

Let's first consider the optimal response of an agent whose perceived cost of the public good is greater than their own willingness to pay. Maintaining the previous assumptions, this agent's optimal response is "zero". This result turns out to be fairly robust to the plausible alternatives to (c) that we will discuss below, and as such, may help to explain the large number of zero responses received to open-ended type questions. The intuition behind this result is that the agent's utility is reduced if the public good is provided and the cost assessed against the agent. The response that adds the least amount to the sum of the benefits (given the usual non-negativity constraint in the open-ended format) is "zero."

Step back for a moment from the benefit-cost criterion that has dominated economic thinking on the incentive structure of the open-ended question and recognize that the simple act of asking an open-ended question is likely to signal to agents that the cost allocation among agents for providing the good is not fixed. Once the agency is prepared to shift the vector of costs facing agents, changing condition (a) above toward increasing the cost to agents having (relatively) high WTP for the good and decreasing it to those who do not, the incentives for agents whose WTP is greater than the initially perceived cost changes substantially. These agents now have to balance the increased probability that the good will be supplied with a high WTP response against the potential upward shift in the cost they will pay if the good is provided. For agents having WTP less than the initially perceived cost, the optimal response is still typically zero.

Since the government rarely if ever uses a pure benefit-cost criterion, it may be plausible for agents to assume that the agency is simply trying to determine what percentage of the relevant population has a WTP higher than the cost, which may or may not be assumed to be known to the agency at the time of the survey. Combined with the potential to reallocate the cost burden, the optimal response of an agent whose WTP is greater than the initially perceived cost is now equal to the cost, while the optimal response of an agent whose WTP is less than the initially perceived cost is still zero.

In all of these cases, the optimal response depends strongly on the agent's perception of the agency's cost of providing the good. The agent should first compare their actual WTP to the expected cost. The optimal response for agents whose WTP is less than the perceived cost, under most plausible uses of the information provided, is zero. These agents should also "protest" in any other way possible if they see this as reducing the likelihood that the good is provided, since if provided they will have to pay more for the good than it is worth to them. This insight puts protest responses in a considerably different light than assumed by many researchers.

The optimal response for an agent whose WTP is greater than expected cost depends upon their belief about how the agency will use their stated WTP. A respondent can generally be seen as maximizing the product of the probability the good is provided times the net benefit that is received if the good is provided. The perceived probability that the good is provided

has been assumed here for a consequential survey to be increasing in the amount provided. For a respondent whose WTP is less than cost, this makes the lowest possible amount the optimal response. This incentive is reinforced, if as likely, the actual cost to be paid is also perceived to be increasing in terms of the amount revealed. However, for an agent whose WTP is greater than expected cost these two incentives conflict and, as such, whether the optimal response should be higher or lower than their WTP indeterminant.[48] To get an unambiguous prediction that an agent's response should be higher than true WTP, it is necessary to assume that the cost to be paid is not influenced by the response provided. This is most likely to happen in the case where the respondent faces a payment vehicle that cannot compel payment from that particular respondent.

Agents do not know cost with certainty. They can and should be expected to formulate priors about the cost and should incorporate any information provided in the survey that they believe is related to cost. This type of behavior would give rise to both starting point bias in iterative bidding games (Boyle et al. 1985) as well as range/placement effects in studies using payments cards to the extent that agents think that the "extra" information provided in these formats is correlated with costs.[49] "Anchoring" on cost or information thought to be associated with cost should be seen as an optimal strategic response even though WTP for the good need have no correlation whatsoever with cost.

On occasion, a variety of different open-ended formats that are said to be incentive compatible are used in a survey context, such as the Becker-DeGroot-Marschak (1964) mechanism, the Vickrey (1961) auction or the procedure proposed by Green et al. (1998).[50] All of these mechanisms elicit a continuous WTP response. There are two things to remember about such mechanisms. First, they cannot get around the Gibbard-Satterthwaite result. Holt (1986) and Karni and Safra (1987) (hereafter HKS) independently showed such mechanisms depend crucially on preferences obeying the expected utility assumption.[51] Many researchers are willing to maintain expected utility and many key economic results on risk are locally robust to most non-expected utility alternatives (Machina 1995).[52] Second, however, when trying to implement any of these mechanisms in a survey context, there is a difficulty that lies much deeper. All of these mechanisms rely on the ability to condition the agent's response on an "exogenous" random element. It can be shown that it is impossible to formulate a simple open-ended matching question that is both informationally and strategically equivalent to an incentive-compatible binary discrete choice question in a survey context. This result is a companion of the HKS theorem. To make the matching question equivalent strategically to the binary discrete choice, the agency has to pre-commit either to a specific cost or to a device that selects the cost independent of the agent's response. This prevents the agency from exploiting the extra information that the agent provides in the matching format but not in the choice format. To get the agent to reveal the true matching answer, the agent

[48] There is every reason to think that there should be considerable heterogeneity with respect to the belief structure concerning how much influence the amount given has on the probability of provision versus the cost to be assessed to agents with characteristics similar to the respondent.

[49] The small number of tests on different payment card amounts (e.g., Rowe et al. 1996) suggests that these effects are not large. One reason for the success of the payment card elicitation format in many instances may be that it actually helps to diffuse an agent's initial cost prior.

[50] Other mechanisms eliciting a continuous-response like the Groves mechanism (Groves 1973) require stronger restrictions on preferences (e.g., quasi-linearity in income) and the possibility of side payments.

[51] It is sometimes thought that the HKS results only apply to goods that are lotteries but Horowitz (2006) shows the HKS results apply even when the good to be valued involves no uncertainty.

[52] See Freeman (1991) for a discussion of environmental risk valuation under non-expected utility preferences.

cannot know the cost. The need for the agent's uncertainty about the cost puts one back in the HKS world where expected utility is required. But more importantly, the need for credible agency pre-commitment not to exploit the extra information contained in the continuous WTP response effectively prevents its being used in a desirable way in a survey context.[53]

7 Concluding remarks

We have argued that serious consideration should be paid to the incentive and informational properties of preference questions. Much of the difficulty with interpreting the apparent anomalies[54] associated with the estimates based on preference survey questions revolves around what we call the face-value dilemma: either agents always truthfully reveal their preferences to the survey question as stated or else they never do. This is a false dilemma. It is easy to demonstrate that a group of agents taken as a whole tend to respond in the direction predicted by theory to the incentives offered in a consequential survey. In some instances the incentives provided by a survey question are incentive compatible for all or most agents and in some instances they are not. A rational economic agent will take the incentive structure of a consequential survey question into account in conjunction with information provided in the survey and beliefs about how that information is likely to be used. We can never rule out, however, the possibility that some agents truthfully respond. We can only say that here, as elsewhere, there are norms that seem to totally or partially override considerations of rationality/self-interest, which is why many individuals would return a lost wallet.

Simple common sense economic models predict large divergences between what agents say they will voluntarily contribute to provide a public good and what they actually contribute. There are now many studies that demonstrate this prediction empirically. The difficulty lies not in the theory or in the experimental demonstrations, but rather in the interpretation that is often placed on these results that *all* survey-based estimates are always unreliable. Further, rather than be taken as evidence that respondents don't have well developed preferences, differences between the estimates obtained using different elicitation formats, if predicted by economic theory, should be taken as evidence supporting the proposition that respondents are taking the scenario posed seriously.

[53] One implication of this result is that studies (e.g., Neil et al. 1994) have tested a survey version of a mechanism like a Vickery auction against an experimental version with real payouts have in actually tested two mechanisms that have very different incentive properties and hence should behave differently. The difference arises because the ability of the agency to act on the information that is provided in the survey case. Since the "purely hypothetical" instruction for the survey treatment that is often used tends to lack credibility, a rational agent might well speculate that giving a high response will increase the likelihood of being asked to participate in some subsequent experiment where real money can be earned.

[54] The term "anomaly" is often loosely used. It is possible to have empirical or experimental results that represent anomalous behavior from the perspective of economic theory, and it is also possible to observe such behavior in surveys. The most interesting anomalies from the perspective of this paper are those that *only* occur in surveys. The first step to take with such an anomaly is to see if it can be observed in settings not involving surveys. A number of anomalies first alleged to be survey-specific have been shown to be easily replicable in experimental contexts and examples readily identifiable in common market transactions. These include preference reversals (Grether and Plott 1979), large divergences between WTP and WTA (Bishop and Heberlein 1990), and part-whole bias (Bateman et al. 1997). In some of these instances, such as the often-noted WTP-WTA divergence, models predicting such divergences consistent with standard neoclassical economic theory have been proposed (e.g., Hanemann 1991). More recently Champ and Bishop (2006) have looked at differences in WTP estimates under different elicitation formats in experiments with monetary payoffs and found differences similar to those typically found in contingent valuation surveys.

Divergences between binary discrete choice and double-bounded formats or between binary discrete choice and open-ended formats are consistent with theory. Optimal response strategies in most cases are fairly simple, and in many instances, such as the zero responses to open-ended type questions, are fairly robust to alternative assumptions made about agent beliefs. In some situations, particular elicitation formats should be avoided altogether, while in others one faces the classic tradeoff of bias versus variance. The researcher should understand the tradeoff being made in the choice of an elicitation format.

Claims about the specific incentive and informational properties of a particular elicitation format should not be made in the abstract. Careful attention must be paid to the type of good being offered, the nature of the payment obligation for the good, and other aspects of the context in which the good is offered in order to properly determine incentive and informational properties. For the binary discrete choice format, the introduction of a new private good turns out to be one of the worst cases for truthful preference revelation, in contrast to previous assertions in the literature to the contrary. The other problematic case is the use of a survey indicator of willingness to voluntarily contribute to estimate the level of actual contributions to a public good. Here neither estimate should approximate the true underlying WTP. One need not cast a binary discrete choice question as a formal referendum to obtain an incentive-compatible question; it is sufficient to structure the question as advice to the government on the issue, a result that should be of use to researchers in areas where referenda are not frequently held.

None of our analysis has relied on agent experience or familiarity with the good. While these may influence the agent's true WTP for the good, they do not influence the incentive properties of question format. Nor have we relied on any notion that agents learn about preferences and update them. Informational and incentive properties of formats do play a role in updating of optimal response strategies. Indeed, it is possible to recast some Bayesian models (e.g., McLeod and Bergland 1999) as Bayesian updating, not with respect to preferences, but rather with respect to determining the optimal strategic response.

A number of elicitation formats commonly used in marketing research are currently attracting considerable attention in environmental valuation, both for the hope that more information can be collected from each agent (than can be collected with the binary discrete choice format) and for the hope that these newer formats will have fewer problems than does a binary discrete choice format. From an incentive perspective, this latter hope is likely to be misplaced with respect to the canonical valuation situation in environmental economics, namely the provision of a pure public good by the government. Problems still exist for quasi-public goods provided by the government but these may be more amenable to the choice experiment format since, in reality, agents tends to face multiple options and the issue of having to determine how to pick a single level of the good applicable to all agents is not present. A different issue not addressed in the paper is that as the number of goods that must be described in a survey increases the time available to describe each good shrinks. For the introduction of new quasi-public or private goods, the multinomial choice format may be close to incentive compatible from the perspective of estimating marginal tradeoffs between attributes, as long as the perceived number of goods that are likely to be provided is sufficiently large and only one of the goods will be consumed. This is because deviations from truthful preference revelation are most likely to impact the scale parameter that drops out of marginal comparisons. This fortunate occurrence is less likely to be true for estimating the total value of a good, since that calculation requires a consistent estimate of the true scale parameter.

In closing, a remark on the term "hypothetical", frequently affixed as an adjective in front of the word "survey", is in order. In a famous and often-cited remark on the early use of surveys for environmental valuation, Scott (1965) bluntly states: "Ask a hypothetical question

and you get a hypothetical answer." "Hypothetical" as used in that quote seems to imply that the responses are to some imaginary *inconsequential* situation, and as such, the responses will have no influence on any relevant decision. From an economic perspective, nothing can be inferred about respondent preferences from the answers to such a question.

The term "hypothetical", however, also means conjecture, counterfactual, and contingent. This is the context typically used by researchers who ask preference questions. It is a term thus potentially consistent with our definition of a *consequential* survey, but an incomplete one because we require the agent to care about the alternatives and to perceive that the agency will take the survey responses into account in its decision making. Our suggestion is to eschew the use of the word *hypothetical* in discussing preference questions, in favor of *consequential* and *inconsequential* to emphasize the conditions requisite for the application of economic theory.

Acknowledgements Support of the U.S. Environmental Protection Agency cooperative agreement R-824698 is gratefully acknowledged. The views expressed here are those of the authors and not necessarily those of the U.S. Environmental Protection Agency. We thank Mark Machina for his considerable help on an earlier version of this paper. The views expressed are those of the authors.

References

Adamowicz W, Louviere J, Williams M (1994) Combining revealed and stated preference measures for valuing environmental amenities. J Environ Econ Manage 26: 271–292

Alberini A (1995) Efficiency vs. bias of willingness-to-pay estimates: bivariate and interval-data models. J Environ Econ Manage 29: 169–180

Alberini A, Kanninen B, Carson RT (1997) Modeling response incentive effects in dichotomous choice contingent valuation data. Land Econ 73: 309–324

Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H (1993) Report of the NOAA panel on contingent valuation. Federal Register 58: 4601–4614

Bateman IJ, Langford IH, Turner RK, Willis KG (1995) Elicitation and truncation effects in contingent valuation studies. Ecol Econ 12: 161–179

Bateman IJ, Munro A, Rhodes B, Starmer C, Sugden R (1997) Does part-whole bias exist?: an experimental investigation. Econ J 107: 322–332

Bateman IJ, Carson RT, Day B, Hanemann WM, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Özdemiroglu E, Pearce DW, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar, Northampton, MA

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single response sequential method. Behav Sci 9: 226–232

Bergstrom T, Blume L, Varian H (1986) On the private provision of public goods. J Public Econ 29: 2–49

Bishop RC, Heberlein TA (1979) Measuring values of extra-market goods: are indirect measures biased. Am J Agric Econ 61: 926–930

Bishop RC, Heberlein TA (1990) The contingent valuation method. In: Johnson RL, Johnson GV (eds) Economic valuation of natural resources: issues, theory and applications. Westview Press, Boulder, CO

Boxall PC, Adamowicz WL, Swait J, Williams M, Louviere J (1996) A comparison of stated preference methods for environmental valuation. Ecol Econ 18: 243–253

Boyle KJ, Bishop RC, Welsh MP (1985) Starting point bias in contingent valuation bidding games. Land Econ 61: 188–194

Boyle KJ, Johnson FR, McCollum D, Desvousges WH, Dunford RW, Hudson SP (1996) Valuing public goods: discrete versus continuous contingent-valuation responses. Land Econ 72: 381–396

Brookshire DS, Ives BC, Schulze WD (1976) The valuation of aesthetic preferences. J Environ Econ Manag 3: 325–346

Burton AC, Carson KS, Chilton SM, Hutchinson WG (2003) An experimental investigation of explanations for inconsistencies in responses to second offers in double referenda. J Environ Econ Manage 46: 472–489

Cameron TA (1988) A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. J Environ Econ Manage 15: 355–379

Cameron TA, Huppert DD (1991) OLS versus ML estimation of non-market resources values with payment card interval data. J Environ Econ Manage 17: 230–246
Cameron TA, Quiggin J (1994) Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire. J Environ Econ Manage 27: 218–234
Carlsson F, Martinsson P (2001) Do hypothetical and actual marginal willingness to pay differ in choice experiments?: application to the valuation of the environment. J Environ Econ Manage 41: 179–192
Carson RT (1985) Three essays on contingent valuation. Unpublished Dissertation, University of California, Berkeley
Carson RT, Steinberg D (1990) Experimental design for discrete choice voter preference surveys. In: 1989 Proceeding of the survey methodology section of the American Statistical Association. American Statistical Association, Washington
Carson RT, Hanemann WM, Mitchell RC (1987) The use of simulated political markets to value public goods. Discussion paper 87–7, Department of Economics, University of California, San Diego
Carson RT, Hanemann WM, Kopp RJ, Krosnick JA, Mitchell RC, Presser S, Ruud PA, Smith VK (1994) Prospective interim lost use value due to DDT and PCB contamination in the Southern California Bight. Report to National Oceanic and Atmospheric Administration
Carson RT, Flores NE, Martin KM, Wright JL (1996) Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Econ 72: 80–99
Carson RT, Groves T, Machina MJ (1997) Stated preference questions: context and optimal response. Paper presented at the National science foundation preference elicitation symposium, University of California, Berkeley
Carson RT, Groves T, List J (2004) Probabilistic influence and supplemental benefits: a field test of the two key assumptions behind using stated preferences. Paper presented at the European association of environmental and resource economists, Budapest, June
Champ PA, Bishop RC (2006) Is willingness to pay for a public good sensitive to the elicitation format?. Land Econ 82: 162–173
Champ PA, Bishop RC, Brown TC, McCollum DW (1997) Using donation mechanism to value nonuse benefits from public goods. J Environ Econ Manage 33: 151–162
Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. J Market Res 19: 281–299
Cronin FJ (1982) Valuing nonmarket goods through contingent markets. Report to U.S. Environmental Protection Agency by Battelle Memorial Institute, Richmond, Washington
Cummings RG, Taylor LO (1998) Does realism matter in contingent valuation surveys?. Land Econ 74: 203–215
Cummings RG, Harrison GW, Rutström EE (1995) Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible. Am Econ Rev 85: 260–266
Cummings RG, Elliott S, Harrison GW, Murphy J (1997) Are hypothetical referenda incentive compatible?. J Polit Econ 105: 609–621
Farmer MC, Randall A (1996) Referendum voting strategies and implications for follow-up open-ended responses. Paper presented at the annual U.S.D.A. W-133 meeting, Jekyll Island, GA
Farquharson R (1969) Theory of voting. Yale University Press, New Haven
Foster V, Bateman IJ, Harley D (1997) Real and hypothetical willingness to pay for environmental preservation: a non-experimental comparison. J Agric Econ 48: 123–138
Freeman A (1991) Indirect methods for valuing changes in environmental risks with non-expected utility preferences. J Risk Uncertain 4: 153–165
Gibbard A (1973) Manipulation of voting schemes: a general result. Econometrica 41: 587–601
Green JR, Laffont JJ (1978) A sampling approach to the free rider problem. In: Sandmo A (ed) Essays in public economics. Lexington Books, Lexington, MA
Green DP, Jacowitz KE, Kahneman D, McFadden DL (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods. Resour Energy Econ 20: 85–116
Greenhalgh C (1986) Research for new product development. In: Worcester RM, Downham J (eds) Consumer market research handbook. 3rd ed. Amsterdam, North-Holland
Grether DM, Plott CR (1979) Economic theory of choice and the preference reversal phenomenon. Am Econ Rev 69: 623–638
Groves T (1973) Incentives in teams. Econometrica 41: 617–631
Groves T, Radner R, Reiter S (eds) (1987) Information, incentives and economic mechanisms: essays in honor of Leonid Hurwicz. University of Minnesota Press, Minneapolis
Haab TC, McConnell KE (1997) Referendum models and negative willingness to pay: alternative solutions. J Environ Econ Manage 32: 251–270

Haab TC, McConnell KE (1998) Referendum models and economic values: theoretical, intuitive and practical bounds on willingness to pay. Land Econ 74: 216–229
Haab TC, McConnell KE (2002) Valuing environmental and natural resources: econometrics of non-market valuation. Edward Elgar Publishers, Northampton
Haab TC, Huang JC, Whitehead JC (1999) Are hypothetical referenda incentive compatible?: a comment. J Polit Econ 107: 186–196
Hanemann WM (1984) Discrete/continuous models of consumer demand. Econometrica 52: 541–561
Hanemann WM (1984) Welfare evaluations in contingent valuation: experiments with discrete responses. Am J Agric Econ 66: 335–379
Hanemann WM (1991) Willingness to pay and willingness to accept: how much can they differ?. Am Econ Rev 81: 635–647
Hanemann WM, Kanninen B (1999) The statistical analysis of discrete-response. In: Bateman I, Willis K (eds) Valuing the environment preferences: theory and practice of the contingent valuation method in the US, EC and developing countries. Oxford University Press, Oxford
Hanemann WM, Loomis J, Kanninen B (1991) Statistical efficiency of double bounded dichotomous choice contingent valuation. Am J Agric Econ 73: 1255–1263
Hausman JA (ed) (1993) Contingent valuation: a critical assessment. Amsterdam, North-Holland
Hausman JA, Ruud P (1987) Specifying and testing econometric models for rank-ordered data. J Econom et 34: 83–104
Hensher DA (1994) Stated preference analysis of travel choice—the state of practice. Transportation 21: 107–133
Herriges JA, Shogren JF (1996) Starting point bias in dichotomous choice valuation with follow-up questioning. J Environ Econ Manage 30: 112–131
Hoehn J, Randall A (1987) A satisfactory benefit cost indicator from contingent valuation. J Environ Econ Manage 14: 226–247
Holt CA (1986) Preference reversals and the independence axiom. Am Econ Rev 76: 508–515
Horowitz JK (2006) The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. Econ Lett 93: 6–11
Hurwicz L (1986) Incentive aspects of decentralization. In: Arrow KJ, Intriligator MD (eds) Handbook of mathematical economics, vol. III.. Amsterdam, North-Holland
Inforsino WJ (1986) Forecasting new product sales from likelihood of purchase ratings (with discussion). Market Sci 5: 372–390
Johannesson M, Liljas B, Johansson PO (1998) An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. Appl Econ 30: 643–647
Johnson FR, Desvousges WH (1997) Estimating stated preferences with rated-pair data: environmental, health and employment effects of energy programs. J Environ Econ Manage 34: 79–99
Kahneman D, Slovic P, Tversky A (1982) Judgment under uncertainty: heuristics and biases. Cambridge University Press, New York
Karni E, Safra Z (1987) Preference reversal and the observability of preferences by experimental methods. Econometrica 55: 675–685
Kriström B (1990) A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. Land Econ 66: 135–139
Kriström B (1997) The practical problems of contingent valuation. In: Kopp RJ, Pommerhene WW, Schwartz N (eds) Determining the value of non-marketed goods. Kluwer, Boston
Lankford RH (1985) Preferences of citizens for public expenditures on elementary and secondary education. J Econo 27: 1–20
Lavrakas PJ, Traugott MW (2000) Election polls, the news media and democracy. Congressional Quarterly Press, Washington
Lohmann S (1994) Information aggregation through costly political action. Am Econ Rev 84: 518–530
Louviere JJ (1994) Conjoint analysis. In: Bagozzi R (ed) Handbook of marketing research. Oxford University Press, Oxford
Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and application. Cambridge University Press, New York
Lunander A (1998) Inducing incentives to understate and to overstate willingness to pay within the open-ended and the dichotomous-choice elicitation format. J Environ Econ Manage 35: 88–102
Lupia A (1994) Shortcuts versus encyclopedias—information and voting behavior in California insurance reform elections. Am Polit Sci Rev 88: 63–76
Machina MJ (1995) Non-expected utility and the robustness of the classical insurance paradigm. Geneva Papers Risk Insur Theory 20: 9–50

Magat WA, Viscusi WK, Huber J (1988) Paired comparison and contingent valuation approaches to morbidity risk valuation. J Environ Econ Manage 15: 395–411
McConnell KE (1990) Models for referendum data: the structure of discrete choice models for contingent valuation. J Environ Econ Manage 18: 19–34
McDowell I, Newell C (1996) Measuring health: a guide to rating scales and questionnaires. 2nd ed. Oxford University Press, New York
McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York
McFadden D (1980) Econometric models for probabilistic choice among products. J Bus 53: S13–S29
McFadden D (1994) Contingent valuation and social choice. Am J Agric Econ 76: 689–708
McFadden D (1999) Rationality for economists. J Risk Uncertain 19: 73–105
McLeod DM, Bergland O (1999) Willingness-to-pay estimates using the double-bounded dichotomous-choice contingent valuation format: a test for validity and precision in a Bayesian framework. Land Econ 75: 115–125
Mitchell RC, Carson RT (1986) The use of contingent valuation data for benefit-cost analysis in water pollution control. Report to the U.S. Environmental Protection Agency
Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Johns Hopkins University Press, Baltimore
Moulin H (1994) Social choice. In: Aumann RJ, Hart S (eds) Handbook of game theory with economic applications. North-Holland, Amsterdam
Neil H, Cummings RG, Ganderton PT, Harrison GW (1994) Hypothetical surveys and real economic commitments. Land Econ 70: 145–154
Payne S (1951) The art of asking questions. Princeton University Press, Princeton
Pearce DW (ed) (2006) Environmental valuation in developed countries: case studies. Edward Elgar, Cheltenham, UK
Pearce DW, Markandya A (1989) Environmental policy benefits: monetary evaluation. Organization for Economic Cooperation and Development, Paris
Pearce DW, Pearce C, Palmer C (eds) (2002) Valuing the environment in developing countries. Edward Elgar, Northampton
Pearce DW, Atkinson G, Mourato S (2006) Cost-benefit analysis and the environment: recent developments. Organization for Economic Cooperation and Development, Paris
Peterson GL, Brown TC (1998) Economic valuation by the method of paired comparison, with emphasis on evaluation of the transitivity axiom. Land Econ 74: 240–261
Popkin S (1991) The reasoning voter. University of Chicago Press, Chicago
Posavac SS (1998) Strategic overbidding in contingent valuation: stated economic value of public goods varies according to consumers' expectations of funding source. J Econ Psychol 19: 205–214
Rabin M (1998) Psychology and economics. J Econ Liter 36: 11–46
Randall A (1994) A difficulty with the travel cost method. Land Econ 70: 88–96
Randall A, Ives BC, Eastman C (1974) Bidding games for valuation of aesthetic environmental improvements. J Environ Econ Manage 1: 132–149
Richer J (1995) Willingness to pay for desert protection. Contemp Econ Policy 13: 93–104
Romer T, Rosenthal H (1978) Political resource allocation, controlled agendas and the status quo. Public Choice 33: 27–43
Rowe RD, Schulze WD, Breffle W (1996) A test for payment card biases. J Environ Econ Manage 31: 178–185
Ruud PA (1983) Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of the distribution in multinomial choice models. Econometrica 51: 225–228
Samuelson PA (1954) The pure theory of public expenditure. Rev Econ Stat 36: 387–389
Satterthwaite M (1975) Strategy-proofness and arrow conditions: existence and correspondence theorems for voting procedures and welfare functions. J Econ Theory 10: 187–217
Scott A (1965) The valuation of game resources: some theoretical aspects. Can Fish Rep 4: 27–47
Seip K, Strand J (1992) Willingness to pay for environmental goods in Norway: a CV study with real payment. Environ Resour Econ 2: 91–106
Smith VK, Desvousges WH (1986) Measuring water quality benefits. Kluwer, Boston
Sudman S, Bradburn NM, Schwarz N (1996) Thinking about answers: the application of cognitive processes to survey methodology. Jossey-Bass Publishers, San Francisco
Sugden R (1999) Alternatives to the neo-classical theory of choice. In: Bateman IJ, Willis KG (eds) Valuing environmental preferences: theory and practice of the contingent valuation method in the US, EU and developing countries. Oxford University Press, New York
Swait J, Louviere JJ (1993) The role of the scale parameter in the estimation and use of generalized extreme value models. J Market Res 30: 305–314

Tversky A, Slovic P, Kahneman D (1990) The causes of preference reversal. Am Econ Rev 80: 204–217
Varian H (1992) Microeconomic analysis. 3rd ed. Norton, New York
Vickery W (1961) Counterspeculation, auctions, and competitive sealed tenders. J Finance 16: 8–37
Vossler CA, Kerkvliet J, Polasky S, Gainutdinova O (2003) Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon. J Econ Behav Organ 51: 261–277