Environmental Valuation

How Much Is the Emperor Wearing?

by

Andrew Stirling

Source: The Ecologist v23 n3 May/June 1993

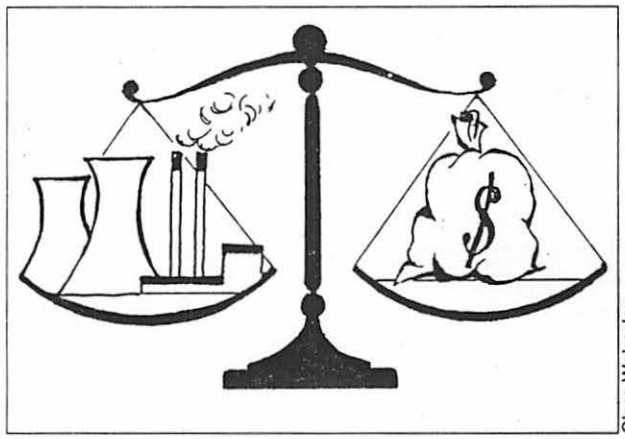
Economists who propose the monetary valuation of environmental effects aspire to produce an objective yardstick for use in policy-making. Yet separate attempts to assign a monetary value to the environmental effects of specific electricity-generating technologies have differed from one another by a factor of as much as 50,000. Discrepancies of this order suggest the existence of fundamental flaws in the basic approach. The complexity of environmental phenomena cannot be expressed by means of a single numerical index, nor can the different perspectives held by various analysts, policy-makers and members of the public ever be reconciled into a single structure of preferences. The adoption of monetary valuation threatens to remove key aspects of environmental decision-making from the sphere of public debate and place it in the hands of a small community of technocrats.

"We are at a point in the evolution of environmental policy at which the economics profession is in a very favourable position to influence the course of policy."

M. L. Cropper and W. E. Oates, 1991

"What in observation is loose and vague is in information deceptive and treacherous."

Francis Bacon, 1621



Clare Walmsley

Modern industrial production causes many forms of environmental damage. There is now strong political pressure to develop credible means to quantify, compare and rank the effects of different technological strategies. Failure to achieve this implies a loss of rationale, and thus of legitimacy, for the environmental regulation of industry. Success will bestow enhanced status upon the favoured community of specialists and an extension of their influence at the expense of competing disciplines. For both policy makers and academics, the stakes are high.

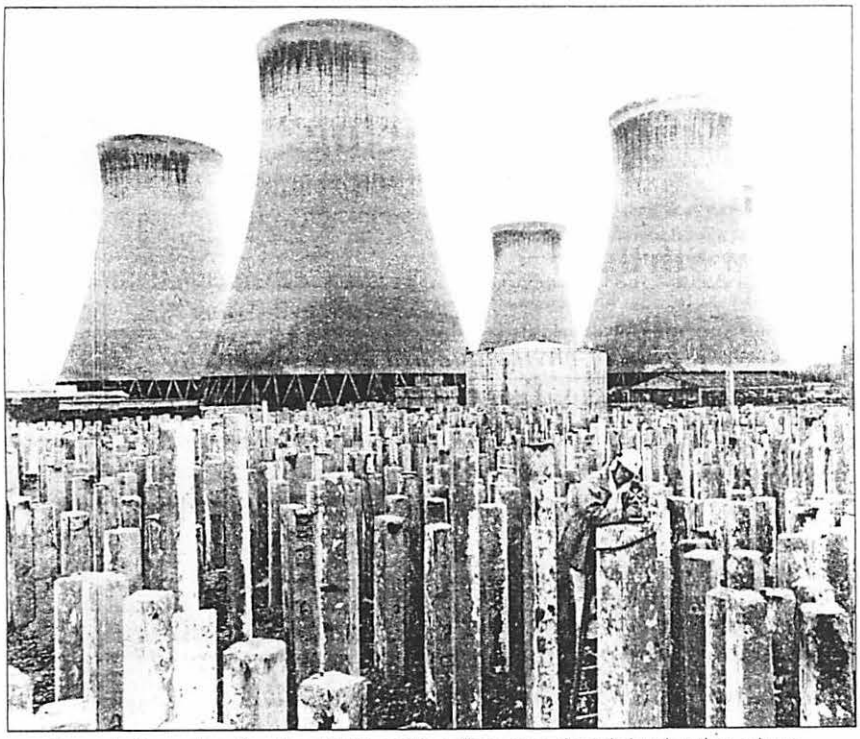
With a rise in the profile of environmental issues over recent years, the pressure has intensified. A wide range of specialists have proposed various approaches: cost-benefit analysis,[1] comparative risk analysis,[2] multi-criteria analysis,[3] decision analysis[4] and environmental impact assessment[5] are among the principal contenders. Efforts have been made to arrange the proliferating number of variants, hybrids and reincarnations as a single palette of techniques, each with its own legitimate domain of application.[6] Nevertheless, a large degree of redundancy persists. Adherence to any one of this set of competing policy tools is often determined more by the disciplinary affiliations of the analyst than by the merits or shortcomings of the method itself.

In the present climate of liberalization and deregulation, neoclassical economics has emerged as an aspiring colonist of this tantalizing but hazardous intellectual territory. Drawing on the concept of social cost introduced by A. C. Pigou,[7] economists have characterized environmental damage as a loss of utility to society as a whole. Monetary value has been proposed as the most appropriate index or yardstick for the measurement of the benefits that are foregone as a consequence of damage to the environment. In the case of those environmental effects which remain unpriced in any market, the problem is to derive monetary values by various analytical means.

Where the economic cost of a reduction in environmental benefits is not included in the price of a commodity, the value of the lost benefits is said to be "external" to that price. Economists view this as a classic case of market failure: their prescription for countering environmental damage is to incorporate these costs into the market price through taxation or other regulatory mechanisms. In this way, they claim, the allocation of resources is optimized with respect both to economic and environmental factors.

It is this intellectual framework which underlies much recent activity in the debate over the regulation of the electricity supply industry. Over the past four years, governmental bodies such as the European Commission, the US Department of Energy and US state regulatory agencies have commissioned a series of studies with the aim of deriving monetary values for the external

Andrew Stirling was the coordinator of Greenpeace International's campaigns against the nuclear industry and nuclear weapons. He is now a doctoral student with the Energy Group of the Science Policy Research Unit at the University of Sussex, funded by the Science and Engineering Research Council.



National Power

Foundation piles for the £700 million flue gas desulphurization plant at Drax power station which will control sulphur dioxide emissions. Are such costs an accurate reflection of the damage caused by pollution ? . . .

environmental costs of the various electricity generating technologies.[8] The results of such studies have already been taken as a basis for legislation relating to the acquisition of new plant. Further and more detailed monetary valuation (or "monetization") studies are reportedly under way, in the US,[9] and for the British Government[10] and the European Commission.[11] Economists are beginning to find themselves in a position of unprecedented influence over environmental policy. But to what extent do the intellectual merits of monetary valuation (which I will henceforth refer to simply as "valuation") justify its current political ascendancy over other attempts to quantify and compare different environmental effects?

Inconsistent Classification

There are a number of problems common to all attempts to quantify environmental effects. The first of these involves classification. Before any scale of measurement is selected (be it money or some other index) the different classes of environmental effect must be accurately characterized in order to avoid duplication or omission. No comparison can be made between studies if they employ different frames of reference.

Yet the classification of environmental effects remains in a chronic state of confusion. In 1983, the OECD noted that few of the published attempts to classify environmental effects were mutually consistent and urged greater care in distinguishing between different classes of harm.[12] For example, "environmental effects" often overlap with "social" or "health effects"; ambiguities in the use of such fundamental terms mask important discrepancies in the scope of different analyses.

Individual classes of effect have variously been defined under the following headings:

- the medium that is physically affected (eg., land, water, air);
- the agent of harm (eg., emissions, residual pollutants);
- the stage in the fuel-cycle or the life-cycle of a facility (eg., mining or decommissioning);
- the form of the risk (eg., catastrophic or routine);
- the manifestation of harm (eg., extinction of species, human illness).

Any one of these organizing principles might be employed to order the set of all environmental effects in its entirety. Conversely, a single effect may legitimately be classified under all. For example, the contamination of a water-course by liquid emissions originating in a catastrophic accident during the construction of an industrial facility may lead to the extinction of a highly localized species.

Where there is confusion over the identification of the various forms of environmental effect, there is a high probability of omitting or duplicating important factors. Efforts to quantify and aggregate environmental externalities are therefore unlikely to yield consistent, comprehensive, or comparable results.

Multidimensional Effects

A second problem common to all quantification techniques is posed by the fact that even the most apparently straightforward environmental effect is an extremely complex phenomenon.[13]

Environmental effects are inherently and irreducibly multidimensional. A single numerical index, such as monetary value, fails to convey important contextual information.[14] Are the data relating to each evaluated technology of equal quality? Are the effects equally familiar to society, and are efforts to mitigate them equally psychologically and socially disruptive? Are there discrepancies between the perceived interests of specialist communities associated with particular technologies and those of society as a whole? Can the consequences of the effects posed by each of the evaluated technologies be avoided, through action taken either before or after their occurrence? Are the effects equally immediate or is, for instance, injury more preponderant with one technology, and disease with another? Is there a direct relationship between the effect and its cause, or does it result from the interplay of complex forces? Is the social distribution of risks correlated with the distribution of associated benefits? Do any of the evaluated technologies pose greater risks to future generations than others? Do all pose the same ratio of occupational to public risks? Or of risks of death to risks of injury and disease? Are the effects associated with the evaluated technologies all equally reversible? How do the range and distribution patterns of the different effects compare? Do certain of the evaluated technologies pose risks of catastrophic effects on a scale unmatched by other technologies? Are the technologies all on identical trajectories in terms of any change in their riskiness? Do their effects differ in the degree of site-specific variation from the estimated norm?[15]

The nature of the risk posed by an individual electricity supply technology depends on each of these factors and many others. As dawns eventually on any child with a "peg and hole"

toy, an ordinary three-dimensional object cannot satisfactorily be characterized in terms of a single parameter such as its "length", its "breadth", or its "depth". In the same way, no one dimension of an effect can adequately convey the totality. Though they may have a superficial appeal, approaches which recognize only a single dimension are as likely seriously to mislead in environmental assessment as they are in everyday measurement.[16] Likewise, the omission of even a single dimension may lead to a seriously deficient understanding of comparative effects.

In the face of these and other difficulties, aspirations "objectively" to quantify, aggregate and compare different classes of environment effect are being superseded in some quarters by more pragmatic aims. One study for the European Commission concludes that aggregation and comparison must necessarily be regarded as political functions, and left to decision-makers rather than specialists.[15] Another EC study, recommends moving away from quantitative cost-benefit analysis toward more qualitative environmental impact assessment.[16] The proponents of monetary valuation, however, proceed as if oblivious to these difficulties.

Pete Addis/Environmental Picture Library

. . . Or do the costs of building new sea-defences to mitigate the damage caused by rising sea-levels give us an reliable measurement of the damage caused by global warming?

Methods of Valuing Environmental Effects

There are many ways of applying a monetary value to environmental effects. Conventionally, three broad approaches are recognized.[17] One unashamedly pragmatic method is to assume that some form of equilibrium pertains, and take the costs of abating, (ie. preventing), harmful effects as a measure of the external environmental costs thereby avoided. Advocates of this approach argue, for example, that the costs of installing flue gas desulphurization equipment in a power station may be taken to represent the costs of the pollution thereby avoided. In such cases any pollution that remains unabated, or even unrecognized, by those in a position of responsibility will remain unaccounted for. Where the object of the valuation exercise is to optimize the allocation of resources for pollution control, taking existing abatement costs as a proxy for damage costs embodies a certain circularity of logic.

A second approach assumes that the cost of mitigating — rather than abating — environmental damage reflects the cost of the damage itself. For example, the depletion of fish stocks by marine pollution can be valued by measuring the increased investment in boats and equipment.[18] However, the estimated costs of mitigating damage tend to cover only those situations where the burden on individual economic actors — and the prospect of relief — are high enough to warrant expenditure. Fishing boat owners may respond to declining catches by selling up rather than increasing investment. Where mitigation costs do not account for damage which is irremediable or too expensive for any affected party to take mitigatory action, they can be taken only as a partial reflection of the total costs of environmental damage.

A third approach seeks to establish empirically the full social costs of the environmental damage itself. Unlike abatement costs or mitigation costs, efforts to determine damage costs at least hold out the prospect of yielding systematically comprehensive values. Unfortunately, this approach is more intractable. There exists a multitude of techniques for assessing damage costs. For instance, a distinction is often drawn between "direct" and "indirect" methods. The "direct" methods include those which:

- establish a value for an "environmental asset", such as a National Park, in terms of the aggregate expenditure on travel (and, sometimes, travel time) by its visitors;
- derive a value by reference to "surrogate" or "hedonic" markets where, for instance, property prices or wages may be seen to take account of the value of environmental or health benefits or disadvantages which are associated with a particular property or job, and;
- conduct "contingent valuation" (CV) by establishing "experimental markets" through responses to questionnaires by sample populations. Respondents state hypothetical monetary values which they would be willing either to pay or to accept in order to secure or forego an environmental benefit, or prevent or tolerate an environmental disbenefit.

An attraction of CV techniques is that they offer the prospect of capturing otherwise intangible benefits, such as those due to the very existence of an environmental amenity (irrespective of its "use"), those associated with the desire to bequeath it to posterity, or those arising from the intention of securing an option on its future use.[19]

"Indirect" damage cost approaches, by contrast, seek to sum the substantive economic costs incurred by all the individual environmental effects. Some of these cost items may be expressed in terms of the market prices for goods and services required in replacement or restoration. These may be assessed by establishing dose-response relationships between the causes and the manifestations of harm. More complex cases, however, must be

ascribed a value by the analyst, which will involve the use of "direct" valuation techniques such as those discussed above.

Unfortunately, the terms "direct" and "indirect" are used in contradictory ways in different areas of the literature, a situation which does not inspire confidence that the framework for the valuation of environmental damages is yet methodical or consistent. But there are more serious grounds for concern over the analytical (and thus regulatory) utility of valuation.

Difficulties with Contingent Valuation

Contingent valuation (CV) is sometimes felt to offer the prospect of deriving values for multidimensional environmental effects which might otherwise prove intractable to valuation. However, where respondents are completely unaware of certain environmental effects, then CV can hardly be said to address these effects. The degree to which different respondents take account of the same effects or dimensions or share the perspective adopted by the analyst remains unexplored. There is thus no way of knowing what fraction of the complete array of dimensions and effects are accounted for.

CV also suffers from more specific difficulties. One important example concerns the discrepancy between the answers to questions focussing on "willingness to pay" and to those focussing on "willingness to accept". Respondents generally cite significantly lower values for what they would be willing to pay in order to secure a particular environmental benefit, than for what they are willing to accept as compensation for its loss. According to one exasperated observer, it took thirteen years of research and sixteen replications before the discrepancy was treated seriously.[20]

Some enthusiasts are still so confident in CV that they prefer to contest the validity of empirical results than to question the theoretical framework itself.[21] But in fields other than economics, such deviations from the behaviour predicted by formal economic theory are regarded as unsurprising. Phenomena dismissed by economists as "cognitive dissonance",[22] are familiar to social psychologists in the guise of concepts such as "loss aversion".[23] It is readily explicable that an individual assigns greater value to attributes with which she or he associates, than to those same attributes prior to any association.

Critical appraisal of CV studies shows that circumstantial factors such as the structure of questionnaires[24] and the demeanour of the questioners may exert a profound influence on results. Far from being passive sources of data, respondents may seek actively to influence the results of studies through various forms of "strategic behaviour". Where rich respondents tend to bid higher values, attributes prized by more affluent communities are likely to be valued more highly in CV. In short, so sensitive is CV to the subjective social and psychological circumstances of respondents and to the contexts of the studies themselves, that some have been led to conclude that "the method becomes the message".

The Partial Scope of Valuation Studies

Values for the external environmental costs of electricity have been derived by the use of each of the techniques discussed above. Noting the shortcomings of the alternatives, different analysts tend to favour different techniques. Some advocate taking damage cost figures such as those generated by contingent valuation or hedonic prices. Others argue instead for taking abatement costs. Some studies draw on a mixture of techniques, thus combining the deficiencies of all. Results generated by different methods diverge to an "unexpectedly" large extent.[25]

The various approaches are often deployed in a somewhat *ad hoc* fashion. There is a tendency to select different techniques for different environmental effects, exacerbating the inconsistent classification of the effects themselves. Important stages of the fuel cycle, or of power station life cycles, (the extraction, transport and storage of raw materials, for example) are routinely excluded.[26] The scope for double counting or omission is compounded by that for misjudgements in the summing of incommensurate valuations. The baroque complexity of the exercise does not make it easy to detect errors. As a result, oversights are committed that would be less likely in more modest (though still daunting) projects such as risk assessment.[27]

The degree to which mitigation costs capture the full scope of a class of environmental effect is also questionable. Certain studies take the cost of liming soil or water as an index of the costs of acid rain, neglecting consequent effects such as those engendered by the extraction of the lime.[28] Others take the cost of improved sea defences on the German North Sea coast as a reflection of the costs of global warming, neglecting the climate effects themselves.[29] Elsewhere, increased investment in irrigation is taken as a representative response to global warming.[30] Such analyses, at best, only partially address those environmental problems they set out to consider.

Rather than making spurious claims to objectivity, policy-makers should acknowledge that calculation is subordinate to judgement

Another crucial issue is the practice of discounting the future. Future benefits are deemed to be financially less valuable than present ones, implying that environmental damage is less harmful the longer it is postponed. For example, the financial benefit of using a superior construction technique to preclude future repairs is not accounted at its nominal value, but at a lower rate to compensate for the delay. The principle can be seen as an economic reflection of the adage that "a bird in the hand is worth two in the bush". It is applied to financial accounts by means of an annual "discount rate", expressed as a percentage. A discount rate of 5 per cent means that that a benefit to be accrued 25 years in the future is assessed at less than 30 per cent of its nominal value.

The chosen discount rate can thus exert a profound influence on the results derived for technologies with different temporal distributions of costs and benefits. Yet the choice of discount rate remains little more than arbitrary and sometimes varies even within a particular study.[31] It is often not clearly declared.[32] Without a knowledge of such factors, the meaningful interpretation of the numerical results of valuation studies is rendered extremely difficult.

Although they may sometimes be aware of at least some of these issues, practitioners of monetary valuation tend to make little effort to acknowledge them. Instead, frequent and prominent use of phrases such as "real costs",[33] "full costs"[34] and "true costs"[35] suggest that valuation results are more systematic and comprehensive than even their authors would admit them to be.

environmental assessment techniques, and adds more of its own. Yet the treatment of uncertainty and variability in results tends to be rudimentary, even compared with the generally lamentable neglect of this problem in other areas of the environmental assessment literature. Valuation results are often expressed with considerable precision.[36] Precision, however, is no guarantee of accuracy.

Theoretical Difficulties in Valuation

The more optimistic proponents of valuation argue that these difficulties will one day be resolved. Such defences fail to explain why valuation is currently so influential amongst policy-makers. Moreover, the theoretical basis for valuation is itself deeply flawed; the extension of the "dubious theology"37 of economics to the arena of environment policy raises profound problems.

The first difficulty concerns the very notion of value. The concept is central to economic theory — in one caricaturist's words, "there is only one value and its name is utility".[38] Yet economists disagree amongst themselves as to its meaning;[39] it would seem that the value of an attribute is inextricably dependent on the context of that attribute and on that of the valuer. This need pose little problem in the case of market or near-market transactions, where the market itself constitutes a common context both for the valuer and the valued. However, if value is "simply a fleeting shadow of wavering contexts, never absolutely existing, and only meaningful in a relative sense",[40] then the concept is ill-suited for use in environment policy, where decisions may have extremely long-term, wide-ranging and profound consequences, far removed from any market.

Second, can the value of environmental attributes properly be expressed in terms of the price society is willing to pay to avoid destroying them? Or does the environment possess some "intrinsic" value in itself, reflecting the benefits secured by non-human organisms? If the latter is the case, then even contingent valuation, which, it is maintained, addresses existence values from the point of view of human respondents, will fail to account for these broader intrinsic values. Although all perspectives are open to the charge of being "sociocentric", valuation differs from other approaches to environmental assessment in that its central index (monetary value) has no meaning whatsoever beyond the confines of certain human societies. The measures employed in other approaches (such as pollutant burdens, mortality, morbidity or toxicity) at least enjoy some substantive physical basis. Although concern for the well-being of non-human life is central to modern environmentalism, this principle seems to lie beyond the analytical scope of valuation.[41]

Third, economists sometimes protest in defence of valuation that a failure to ascribe monetary values to environmental attributes implies that such attributes are of infinite value.[42] Elsewhere it is claimed that a failure to "monetize" implies the ascription of no value at all.[43] Ignoring any contradiction, if either argument were valid, it would do no more than highlight a phenomenon at the heart of human experience. It is obvious that the refusal of a parent to ascribe a monetary value to their child need no more be seen to indicate a zero valuation than it does an infinite valuation. Certain forms of value are simply beyond price. Far from being an inconsistency, this offers an everyday

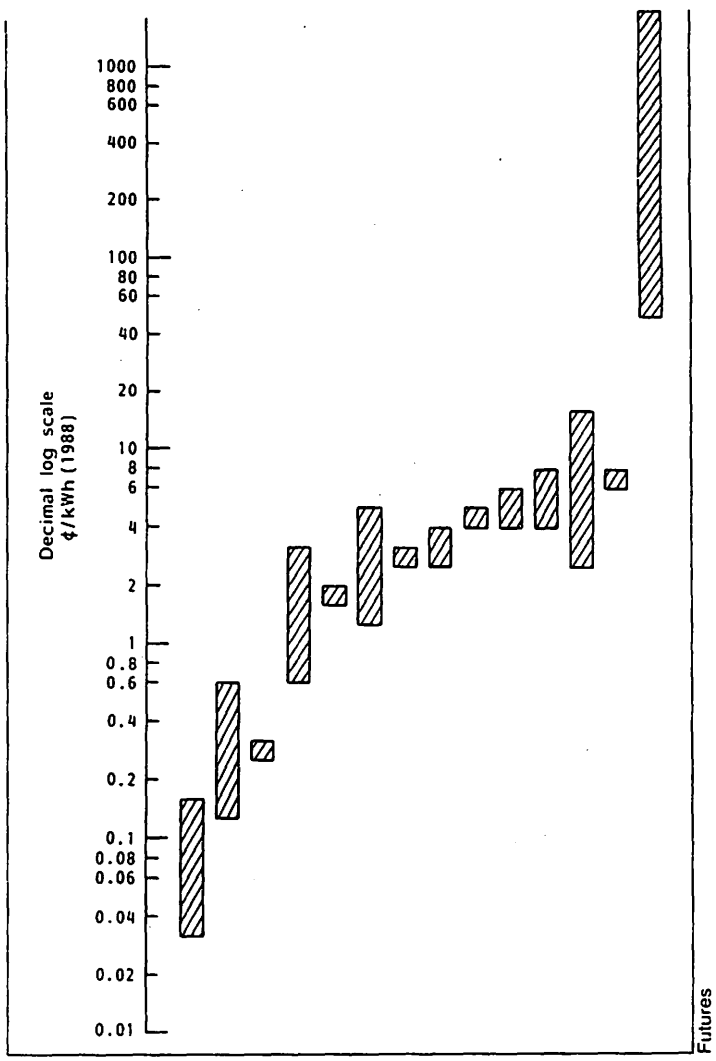


Figure 1. *Range of selected estimates of environmental costs of coal-fired electricity generation.*

illustration of the difficulty of characterizing an multi-dimensional whole by means of a one-dimensional index. Although there undoubtedly exist monetary components to the value of many environmental attributes, it is both naive and perilous to take these as a reflection of the totality. Attempting to encapsulate environmental quality in a monetary value is like trying to measure the width of a temperature, or divine the velocity of love.

The Practical Results

The discussion so far has highlighted grounds for concern over the valuation project. To what extent are these borne out in practical results? A large number of studies look at the external costs of coal-fired electricity — many more than for any other power-generating technology. The lower and upper bounds to the published range of results of these studies differ by a factor of more than fifty thousand (see Figure 1[44]). Although different analysts in any discipline will employ different frames of reference, use different data, and adopt different assumptions, or methodologies, a range of variation exceeding four orders of magnitude is difficult to explain in these terms. At the very least, the scale of the disagreement suggests that the accuracy of valuation does not match the precision with which individual authors express their results.

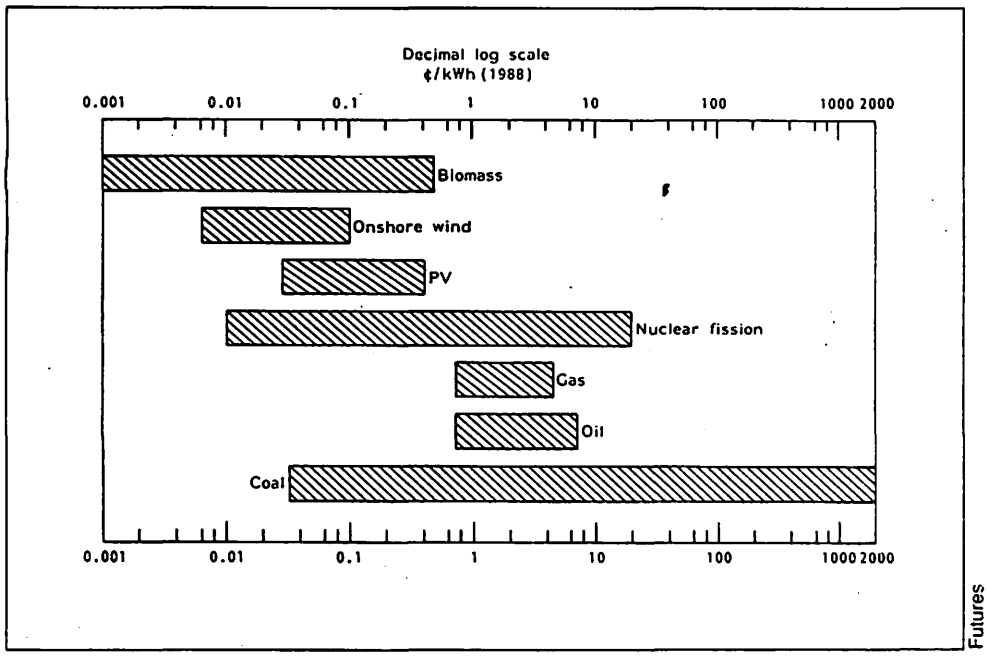


Figure 2. *Overlap between ranges of reported environmental externality estimates for selected electricity supply technologies.*

The magnitude of this range of variation has two serious implications. The first relates to the significant overlap between the ranges of external costs attributed to different technologies (see Figure 2). This overlap is sufficient to accommodate a multitude of different ways of ranking the technologies. If valuation is not accurate enough to provide a basis for confident discrimination between competing technical options, then its policy utility is seriously undermined.

The second concerns the reasons why the figures for externalities take the values they do. In figure one, although the total range of all the estimates is very wide, the majority of published results cluster around the established market price of electricity (Figure 1). This is precisely the range of values which might most readily be incorporated into market prices as an environmental "tax". Values much higher than this would not be directly usable, since the effect of increasing electricity prices by factor ten might be thought to be prohibitive. Values much lower than this would be too small to encourage tangible behavioural changes.

Those seeking a reason for what appears to be a convenient correlation between valuation results and market prices are left with an invidious choice. Are these "accurate" results which suggest mysterious natural mechanisms linking the physical world and the market economy? If so, it is curious that this phenomenon is not often cited as an endorsement of valuation. Alternatively, are there powerful social mechanisms acting on valuation researchers which ensure that valuation results tend to lie in the range most useful in policy-making? Failing either of these, this strange correlation could, of course, simply be coincidence.

The Social Implications of Valuation

The impetus for the development of valuation and other approaches to environmental assessment lies in a belated recognition that orthodox economic analysis since the Industrial Revolution has failed to take adequate account of the environment. Rather than acknowledging that this failure may reflect shortcomings in the discipline of economics, those who advocate valuation have taken the opposite route. Their central thesis is that the environment has been neglected because the concepts and rationale of neoclassical economics have not been applied extensively enough. The notion of monetary value, they say, should be extended from the domain of the economy to the domain of the environment.

Aspirations to reduce complex problems and relationships to simple numerical terms are far from new. Analysts have always been prone to "confuse things that are countable with things that count".[45] Seduced by the facility of calculation, they tend, in Landsberg's words, "to become fascinated with the numbers that emerge and to look at them as real-world, mutually independent variables rather than as the end result of a large number and variety of non-verifiable hypotheses and sheer guesses".[46] Such a tendency is manifest among the more credulous proponents of environmental valuation

Monetary valuation is *scientistic* in the sense that it relies for its authority on the willingness of policy makers and the general public to accept the validity of ostensibly precise numerical results as adequate expression of complex; context- dependent and multidimensional qualitative issues. It is also *technocratic*, in the sense that it delegates important political judgements to specialists to an extent greater than other techniques and so is even less transparent to informed public scrutiny and consent. Such defects are also shown by risk assessment and energy-demand forecasting. Historically, each has proven highly vulnerable to manipulation by powerful institutional interests, such as those of the nuclear industry. Just when environmentalists have largely succeeded in discrediting such procedures, well motivated environmental economists risk presenting a new and even more attractive opportunity to industrial special pleading.

The alternative to valuation lies in acknowledging the fundamentally multidimensional character of environmental effects. The complexities of nature and human society are better represented by a number of decision making criteria. Such criteria are far more effectively identified and prioritized through wide political debate, than by small communities of specialists with minority conceptions and interests. Rather than making spurious claims to objectivity, policy-makers should acknowledge that calculation is subordinate to judgement — that the selection and ranking of environmental criteria are inevitably subjective. Although a plural society is unlikely ever to reach consensus over the final choice of criteria, such an admission would at least provide a basis for more accessible political debate.

This article is adapted from the text of a longer paper, "Regulating the Electricity Supply Industry by Valuing Environmental Effects", published in *Futures*, Volume 24, Number 10, December 1992, The author is grateful for helpful comments from Jim Skea, Gordon MacKerron, Nick Von Tunzelmann, Catherine Mitchell and Topsy Jewell and for inspiration from Roy Harper and the conker tree. The numerous flaws are of his own making.

References

1. Eg. Pearce, D.W. and Nash, C.A., *The Social Appraisal of Projects: a Text in Cost-Benefit Analysis*, 1981; Cropper, M.L. and Oates, W.E., *Environmental Economics: a Survey, Resources for the Future*, Discussion Paper QE90-12-REV, 1991.
2. Eg. A.V. Cohen, A.V. and Pritchard, D.K., *Comparative Risks of Electricity Production: A Critical Survey of the Literature*, UK Health and Safety Executive, 1980; Chadwick, M.J. et al., (eds.) *Comparative Environmental and Health Effects of Different Energy Systems for Electricity Generation*, Key Issues Paper No. 3, presented to the Senior Expert Symposium on Electricity and Environment", convened by the International Atomic Energy Agency (IAEA) et al., Helsinki, May, 1991. .
3. Eg. Keeney, R.L. and Raiffa, H., *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Wiley, 1976.
4. Eg: Von Winterfeldt, D. and Edwards, W., Decision Analysis and Behavioural Research, Cambridge, 1986.
5. Wathern, P., (ed.), Environmental Impact Assessment: Theory and Practice, Unwin Hyman, 1988; Lee, N., Environmental Impact Assessment: a Training Guide, report prepared for DG XI of the European Commission by the EIA Centre, Occasional Paper 18, University of Manchester, 1989.
6. Eg. Pearce, D.W. and Markandya, A., *Environmental Policy Benefits: Monetary Valuation*, OECD, 1989.
7. Pigou, A., *Economics of Welfare*, MacMillan, 1920.
8. Eg: Hohmeyer, O., *Social Costs of Energy Consumption: External Effects of Electricity Generation in the Federal Republic of Germany*, prepared for DG XII of the European Commission by the Fraunhofer Institut für Systemtechnik und Innovationforschung, Springer, 1988; Ottinger, R.L. et al., Environmental Costs of Electricity, prepared for the NYSERDA and US DOE by Pace University Center for Environmental Legal Studies, Oceana, 1990.
9. Cited in Desvousges, W.H. et al, *Accounting for Externality Costs in Electric Planning in Wisconsin: Final Report*, prepared for the Task Force on Externality Costing by the Research Triangle Institute, November, 1991.
10. House of Commons Select Committee on Energy, *Renewable Energy*, 1991-2 Session, Fourth Report, HMSO, Vol. III, London, 1992, p.123.
11. Commission of the European Communities, "Memorandum to British Parliamentary Select Committee on Energy", January 1992, in House of Commons, op. cit.10.
12. Organisation for Economic Co-operation and Development, "Environmental Effects of Energy Systems: the OECD Compass Project", OECD, 1983.
13. Proponents of comparative risk assessment have been forced to acknowledge this after two decades of debate. Chadwick, M., op. cit. 2, 1991: p.7. Eg. Holdren, J.P. et al., "Energy: Calculating the Risks", *Science*, 204, 1979; Organisation for Economic Co-operation and Development, "Environmental Costs of Electricity Production", OECD, 1985: p.130.
14. This set of questions is neither comprehensive in scope nor exhaustive in depth. See Stirling, A., *Technology Choice for Electricity Supply*, unpublished D.Phil. thesis, University of Sussex, 1992.
15. Rowe W.D. and Oterson, P., "Assessment of Comparative and Non-Comparative Factors in Alternative Energy Systems", European Commission, Brussels, 1983.
16. Lee, op. cit. 5, 1989.
17. Eg. Pearce, D.W., Markandya, A. and Barbier, E., *Blueprint for a Green Economy*, Earthscan, London, 1989; Pearce, D. and Markandya, A., op. cit., 1989: 12; Johansson, P., "Valuing Environmental Damage", *Oxford Review of Economic Policy*, 6, 1, 1990; Cropper, M. and Oates, W., op. cit. 1, 1991.
18. Cropper, M. and Oates, W., op. cit. 1.
19. "Existence" values are sometimes distinguished from (eg. Pearce et al., op. cit. 17, p.61) and sometimes taken to subsume (Johansson, op. cit. 17, p.38), "option" values and "bequest" values (Krutilla, J.A., "Conservation Reconsidered", *American Economic Review*, 57, 1967). Interestingly, different opinions are expressed at different times by the same authorities (compare Pearce, D. et al., op. cit. 17, p.61 and Pearce, D. and Markandya, A., op. cit. 6).
20. Heberlein, T.A. "Economics and Social Psychology in Amenity Valuation", in Peterson G.L. et al. (eds.), Amenity Resource Valuation: *Integrating Economics with Other Disciplines*, Venture, Philadelphia, 1988.
21. Reported in Pearce, D. and Markandya, op. cit. 6, 1989: page 14.
22. Ibid.; and Cropper, M., and Oates, W., op. cit.1, 1991: page 65.
23. Schroeder, H.W. and Dwyer, J.F., "Overview: Gains, Losses and Contingent Valuation", in Peterson, G., et al., op. cit. 20.
24. Brown, T.C. and Slovic, P. "Effects of Context on Economic Measures of Value", in Peterson, G., et al., op. cit. 20.
25. Johansson, op. cit. 20, 1990.
26. For example, compare Ottinger et al., op. cit. 8, with Hohmeyer, op. cit. 8.
27. For instance, Ottinger, R. et al., ibid., make an arithmetic error of factor in computing nuclear accident risks. Both Ottinger et al. and Hohmeyer, ibid. entirely omit consideration of the aesthetic impacts of wind power.
28. Ottinger, R., et al., ibid.
29. Hohmeyer, O., op. cit. 8, .
30. Cropper, M., and Oates, W., op. cit. 1.
31. Ottinger, R. et al., op. cit. 8 .
32. Eg. O. Hohmeyer, "Latest Results of the International Discussion on the Social Costs of Energy: H*ow does Wind Compare Today?*, paper presented to the 1990 European Wind Energy Conference, Madrid, October 1990.
33. Eg. Hubbard, H. M., "The Real Cost of Energy", *Scientific American*, 264, 4, 1991; Hirata, G.N., Takahashi, P.K., *Assessing the Real Cost of Energy*, prepared for the Conference of the International Association for Energy Economics, July 1991.
34. Eg: Hohmeyer, O., "Renewables and the Full Costs of Energy", *Energy Policy*, Vol.20, No.4, April 1992.
35. Ottinger, R., "Getting at the True Cost of Electric Power", *Electricity Journal*, July 1990.
36. Voss et al, "Externe Kosten der Stromerzeugnung, Studie im Auftrag der VDEW", Frankfurt, VDEW, 1989; and Ottinger, R. et al. op. cit. 8; both give some results to three significant figures; Hohmeyer, O., op. cit. 8, and op. cit. 32, presents results in four significant figures.
37. Brown, P.G., "Climate Change and the Planetary Trust", *Energy Policy*, March 1992.
38. Boulding, K.E., and Lundstedt, S.B., "Value Concepts and Justifications", in Peterson, G., et al. (eds.), op. cit. 20, 1988.
39. Heilbroner, R., "The Problem of Value", in Heilbroner (ed.), *Behind the Veil of Economics: Essays in the Worldly Philosophy*, Norton, New York, 1988.
40. Peterson, G.L., Driver, B.L., Gregory, R., "Let Us Hear the Conclusion of the Matter", in Peterson G., et al. (eds), op. cit. 20.
41. This is well illustrated by Johansson, op. cit. 17, p.34, who states that "If a forest which is used as a recreation area is cleared, those owning the forest will gain while those visiting the area or those concerned about an endangered species living there will lose". In this statement, (i) no value is attached to the preferences of non-human species, (ii) value is attached only to individual organisms where they are of endangered species, and (iii) it is implied that economic values may in any case only be attributed in respect of those human beings who happen to be "concerned".
42. Although acknowledging that use of the term "priceless" need not be taken to imply infinite valuation, Pearce et al., op. cit. 17, rebut this latter interpretation as a straw man.
43. Ottinger, R. et al., op. cit. 8.
44. Sources for these estimates may be found in Stirling, A., "Regulating the Electricity Supply Industry by Valuing Environmental Effects", *Futures*, December 1992; or by writing to the editors of *The Ecologist*.
45. Holdren, "Energy Hazards: What to Measure, What to Compare", *Technology Review*, February 1982.
46. Landsberg, H. H., "Commentary on Paper by A.M. Perry", in W.C. Clark (ed.), *Carbon Dioxide Review*, Oxford, 1982.

FINAL REPORT

Assessing the Potential for A Total Valuation Study of Colorado River Resources

July 10, 1991

Prepared for:

Glen Canyon Environmental Studies
125 East Birch Street, Suite 307
Flagstaff, AZ 86001

Prepared by:

HBRS, Inc.
585 Science Drive, Suite A
Madison, WI 53711-1060
(608) 231-1011

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The Glen Canyon Environmental Studies (GCES) is a program of research focusing on how operation of Glen Canyon Dam on the Colorado River in northern Arizona affects the downstream resources along the Colorado River. The river flows freely for nearly 300 miles below the dam, mostly within Grand Canyon National Park. In addition to storing water in Lake Powell for eventual delivery to the Lower Colorado River Basin States and Mexico, Glen Canyon is a major resource for the generation of electric power. Within water storage and delivery schedules designed to meet commitments to the Lower Basin States under the Colorado River Compact, the dam and power plant are currently operated to generate sufficient power revenues to repay project costs.

Several environmental and recreational resources are potentially affected by dam operations. The viability of riparian ecosystems of the Grand Canyon, the status of cultural resources of historic and pre-historic Indians, and recreation depend on beaches along the river corridor that are eroding away. Dam operations may affect the rate of erosion. Fish populations, including a population of the endangered humpback chub and other native fishes, as well as non-native fish such as trout, are affected as well. Other environmental resources may also be affected by the dam. Dam operations also influence whitewater boating and recreational fishing. GCES was assigned to determine whether dam operations are having adverse effects on such resources and, if so, what might be done to reduce such effects.

Much of the research under GCES has necessarily involved the physical and biological sciences in wide-ranging efforts to understand how the river acquires, moves, and loses sediment and how the ecosystems along the river function. In addition, economic research continues to be a integral component of the research program. To date, economic research has had two foci. First, economic tools are being employed to quantify, in monetary terms, the effects of dam operations on the quality of river recreation (Bishop et al. 1987, Department of Interior 1988). Secondly, if dam operating criteria were altered to reduce adverse environmental or recreational effects downstream, this would almost certainly affect the economic value of the electricity generated at the dam. Hence, studies are in progress to understand the effects of dam management alternatives on the value of power produced at Glen Canyon Dam.

Recreation and power values are use values, since they stem from the direct use of river resources to produce electrical and recreational services that have value to individual citizens. Environmental economists, beginning in the mid-1960s, have increasingly wondered whether the economic values of environmental resources should be limited to use values. Grand Canyon National Park makes an excellent example. Are the economic values of the park limited to the values held by those who actually go there? It seems plausible that others who are not currently engaging in recreation in the canyon might place an economic value on the preservation of the park's resources. For example, they might place some value on preserving environmental resources for future generations. Likewise they may place a value on preserving environmental resources to maintain the option of using the park's recreational opportunities in the future. Such values are often called non-use values. To try to incorporate the broader values that may not be motivated by use, environmental economists have developed a theory of total value, in which total value consists of both use and non-use values.

The research reported here was commissioned to help those responsible for planning GCES research to decide whether or not economic studies under GCES should be expanded to better account for the total values--including non-use values--of the resources under study. On the one hand, given that Grand Canyon resources are affected by Glen Canyon Dam operations, some would argue that use values alone are inadequate. According to this argument, members of the broader public, and not just recreationists and power users, place a value on Colorado River resources, and these non-use values must be counted or the economic analysis under GCES could be woefully incomplete. On the other hand, others would argue that non-use values are still too new, too poorly understood and too difficult to accurately measure to be useful in making decisions about future operating criteria for the dam. According to the latter view, the non-use values that might ideally be incorporated through application of total value concepts would be better left for consideration in non-economic, more descriptive, more qualitative terms.

Questions about possible research on the total values of the resources downstream from Glen Canyon Dam are being raised at a particularly appropriate time. An Environmental Impact Statement on the operations of the dam is currently in preparation. The law requires consideration of the economic implications of alternatives considered as part of the EIS process. As we shall see in the coming pages, the economic theory and empirical measurement

techniques needed to incorporate non-use values in resource valuation studies have evolved rapidly during the last decade. As a result, non-use values are beginning to find their way into practical applications. Perhaps the most important example of this trend is to be found in the rules for assessing the damages to natural resources from spills of oil and toxics under CERCLA and the Clean Water Act (U.S. Department of the Interior, 1987), damages for which responsible parties are to be held liable. Non-use values were allowed under the original rules and, when those rules were struck down in a U.S. Court of Appeals decision in 1989, the court did so in a way that strengthened the role of non-use values rather than weakening it. Given the increasing attention to total valuation in such instances, whether economic analysis for the Glen Canyon Dam EIS should include them becomes a very important issue.

At its core, the current debate over inclusion of non-use values as a fully fledged component of total value is a debate over their scientific credibility. In the first two sections of this report, we shall begin to assess the potential credibility of non-use value estimates for resources of concern to GCES by summarizing the current status of economic research in the area. Both theoretical and empirical issues are addressed.

Section 1 focuses on the theoretical side of the literature. In order to estimate natural resource values, a conceptual framework is needed that thoroughly integrates use and non-use values into a total valuation framework. In addition to defining key valuation concepts and exploring the relationships between different value concepts, Section 1 will also address a number of theoretical issues having special relevance to non-use values. While additional work is needed on the theoretical level, total values, including non-use values, appear to be well on their way to being established in theory. Several issues for possible empirical research under GCES will be identified through consideration of the theoretical issues.

Section 2 addresses issues of measurement. Regardless of the theoretical validity of total valuation concepts, if they cannot be measured with an acceptable degree of accuracy, they should not weigh heavily in policy decisions, at least in a quantitative sense. It is generally acknowledged that full measurement of total values when non-use values are present must depend on contingent valuation. Thus, Section 2 will review results from past contingent valuation studies that have attempted to measure at least some of the non-use aspects of total values and available evidence regarding the potential accuracy of the technique. While a certain

degree of skepticism is warranted in any new area of research, concerns about contingent non-use values seem to us to be exaggerated.

The literature on theory and measurement then serves as a basis for the third section of this report, which presents a preliminary research prospectus for a total valuation study under GCES. What would be the major components of such a study? Based on our review of theory and past studies, what issues should be explicitly addressed? What potential pitfalls might be encountered and how might they be avoided? The proposed research described in Section 3 would proceed in several steps, each designed to answer critical theoretical or empirical questions. At the end of each step, a review of results would be made and a decision reached as to whether to proceed to the next step. After each step, the research plan would need to be modified and further developed in light of comments from experts and the concerns and needs of GCES agencies and the public.

To further evaluate the potential scientific credibility of total valuation studies of Colorado River resources, the first draft of this report was submitted to a panel of leading environmental economists. The panel was composed of Professor Ann Fisher of Pennsylvania State University, Professor A. Myrick Freeman, III, of Bowdoin College and Resources for the Future, Inc., Professor Alan Randall of the Ohio State University, and Professor V. Kerry Smith of North Carolina State University. The panel met for an introductory workshop in December 1990 and again for an intensive review of the first draft of this report in March 1991. In addition, panel members submitted brief written comments in which they were asked to address three questions. First, are total value concepts including non-use values applicable and potentially significant for the resource being studied under GCES? Second, should efforts be made to measure total values as part of GCES and the Glen Canyon Dam Environmental Impact Assessment process? Third, can total values for Colorado River resources be estimated in a scientifically credible way? In reviewing the draft report both orally and in writing, the panel made many useful comments and suggestions. To the extent that it was practically possible, those suggestions have been incorporated in the revisions leading to this final draft of the report. In addition, the written comments of the panelists are included as an appendix.

Because this report is concerned with scientific credibility, it is addressed to other economists and sometimes is necessarily rather technical. Those without the interest or background to wade through the technical material in the coming pages may find a non-technical discussion of the issues we discuss and the conclusions we reach in the Executive Summary.

SECTION 1: ECONOMIC THEORY OF TOTAL VALUATION

The goal of this section is to draw upon the literature as needed to develop a theory of total value for Colorado River resources "from the ground up," so to speak. That is, the analysis will begin with a straightforward case in which total value is comprised solely of market use values, using electric power as a case example. Then, non-market use value will be added to account for recreational uses of the resources. Next, existence value will be added to the theoretical model. Toward the end of Section 1, uncertainty will be introduced so that option value can be addressed.

Formal Definition of Use Values for Market Goods

The traditional analysis starts by characterizing the actions of an individual as that of choosing a utility-maximizing set of marketed goods subject to a budget constraint. More formally the problem is characterized as:

$$\text{Max } U(X) \qquad \text{subject to } p\,X \leq I$$

where $U(\cdot)$ is a utility function representing the preferences of the individual, X is a vector of goods that can be purchased, p is a vector of prices for these goods, and I is the individual's income. We make the usual assumptions about the characteristics of the utility function. The solution to this problem, X^*, will generally be a function of p and I. Substitution of the market demands, $X^*(p,I)$, for X in the utility function defines the indirect utility function, $V(p,I)$. The dual associated with the utility maximization problem can be written:

$$\text{Min } pX \quad \text{s.t. } U(X) \geq \overline{U}$$

The solution to this problem, X^c, will be a function of prices and the reference utility level. Substitution of the optimal choice, $X^c(p,U)$, into the objective function yields the expenditure function e(p,U). The expenditure function gives the minimum cost of obtaining a reference level of utility at any price level.

Now suppose that our representative consumer buys electricity from a utility that has an allocation of Glen Canyon power. Suppose further that an alternative set of dam operating criteria are being considered that would increase the price of power faced by this consumer so that her price vector would shift from p, under the status quo operating criteria, to p'. Note that if by changing the dam operating criteria, the prices of goods and services other than electric power would be affected through rising energy costs, this would also be depicted in p'.

The indirect utility function and the expenditure function are important because they provide the conceptual framework used to derive dollar based measures of the impacts of changes in dam operations. First using the indirect utility function, two measures of the loss sustained by this consumer can be defined. Her willingness to pay (equivalent variation) to avoid the loss due to the change in dam operating criteria would then be represented by E in:

$$V(p', I) = V(p, I-E)$$

Compensating variation, reflecting compensation required to offset the adverse effects of the change in operating criteria, can likewise be represented as C in:

$$V(p', I+C) = V(p, I)$$

From a purely theoretic view, both C and E represent valid measures of value. However, as will be seen in Section 2, measurement problems appear to be more severe for values measured as willingness-to-accept compensation. Consequently, we shall develop the analysis only in terms of willingness to pay, with two caveats. First, parallel willingness-to-accept measures can always be constructed and are potential, and equally valid, alternative welfare measures. Second, depending on whether we are valuing a loss or a gain relative to the status

quo, willingness to pay will sometimes be defined as the compensating variation and sometimes as the equivalent variation.

Willingness to pay can, of course, also be measured using the expenditure function. In this way,

$$E = e(p', v(p',I)) - e(p, v(p',I))$$

or, using the property that:

$$\frac{\partial e(p,U)}{\partial p_i} = X_i^c(p,U)$$

$$E = \int_{p}^{p'} X^c(\tau, v(p',I))\, d\tau$$

In this rather traditional presentation of valuation theory for a price change, one measure of value is simply the maximum amount of income the person would sacrifice to avoid losing access to the electricity at the lower price. This amount can be represented in terms of utility functions, expenditure functions, or areas under compensated demand functions. It is the ability to translate the value measure into areas under compensated demand functions that has been very important. While the compensated demand functions are unobservable, there is a close relationship between ordinary demand functions and compensated demand functions. A standard approach for measuring economic impacts associated with marketed goods then is to estimate the relevant demand function using observations of actual transactions and then use the demand functions to recover an estimate of the economic impact of the change being considered.

This standard approach, however, rests on the implicit assumption that the consumer in question is only potentially affected by the change in dam operations through changes in the

price of electricity and possibly the prices of other goods and services acquired through the market system. Next we consider the case where the consumer would be affected through the impacts of the change on the availability of a non-market good. As an example we will consider changes in the conditions under which she engages in Colorado River recreation.

Use Values of Non-Market Goods

Let us suppose that the change in dam operations would improve the conditions under which whitewater recreation is conducted. To include this potential impact in the analysis, the utility function is modified so that utility is a function of market goods and Q, where Q is taken to represent the environmental attributes that contribute to the quality of a rafting trip in the Grand Canyon. The price vector, p, and market consumption vector, X, are now taken to include a basket of market commodities required to make a raft trip using an appropriate household production technology, defined in the usual way. Modifying the utility function in this way means that the indirect utility and expenditure functions will also depend on the level of Q.

Assume that under the present dam operations (taken to be the status quo), recreational quality is summarized in a vector, Q, and that the quality would improve under the possible change in operating criteria to Q′ (i.e., $U(X,Q) < U(X',Q')$, where X and X′ are the optimal vectors of market goods consumed under Q and Q′, respectively). While little needs to be said about quality vectors at the present level of abstraction, they obviously include a very important set of variables representing what Smith, in his comments in the appendix, calls the "service flows" from the resources in question. Establishing which variables to include in the quality vectors will be an important part of any total valuation research under GCES.

If the whitewater rafter in this example does not purchase any market goods whose prices might be affected by a change in dam operations total value will consist solely of non-market use values. This total value can be measured as willingness to pay for the improvement in quality. This willingness to pay is given by C, which can be defined as:

$$V(p, I, Q) = V(p, I - C, Q')$$
$$C = e(p, Q, V(p, I, Q)) - e(p, Q', V(p, I, Q)).$$

Notice that in this case, we have ignored the effect of the change in operations on the price of electricity. However, accounting for the fact that our consumer might use both electricity from Glen Canyon Dam and Colorado River recreational resources presents no theoretical difficulties. Suppose that the change in operating criteria would result in an increase in the price of electricity (and perhaps other marketed goods) as well as an improvement in recreational quality. Suppose further that:

$$V(p', I, Q') > V(p, I, Q).$$

Then, willingness to pay would be defined by:

$$V(p, I, Q) = V(p', I - C, Q').$$

If the consumer would be worse off, on net, with the combined price and quality changes, the comparable measure of equivalent variation would apply. There are no difficulties with integrating market and non-market use values within the same economic framework.

An Initial Look at Total Values, Including Existence Values

The foregoing discussion began with use values associated with market goods and then broadened the focus of economic analysis to deal with use values of non-market goods. The next step involves expanding the theory to include total values under certainty, including existence values. On a relatively abstract level, at least, total values can be defined using a simple extension of the previous results. The extension is to expand the vector Q to include all environmental variables for the section of the Colorado River under discussion that could affect the individual's utility level. As above, some elements of Q may affect the quality of a raft trip, for example, the availability of beaches for camping. Other elements, such as the condition of the population of humpback chubs, might not affect recreational quality but would still be

included in Q if they have an effect on the individual's utility. We assume that the elements of Q are defined in such a way that an increase in any element of Q increases utility or at least leaves it unchanged. Finally, we will define a price vector for market goods, $p^{"}$, such that the consumer would not make a raft trip or otherwise use the non-market resources of the canyon. Under our definitions, existence value would be present when $Q' > Q$ implies that $V(p^{"},I,Q) < V(p^{"},I,Q')$. The same logic used in defining willingness to pay before can be easily extended to formulate one definition of existence value, EV_0, as:

$$V(p^{"}, I, Q) = V(p^{"}, I - EV_0, Q').$$

Defined in this manner, EV_0 is a conditional existence value in the sense defined by Boyle and Bishop (1987). It is conditional in that all non-market uses of the resource are constrained to be zero by definition of the price vector. However, we do not assume that $p^{"}$ is such that market uses such as consumption of electricity are zero. It seems plausible that the marginal utility of electricity consumption is not dependent on any of the elements of Q.

For a variety of reasons, a large majority of people, even from Arizona and neighboring states, never have direct contact with the riparian ecosystems and other resources of the Colorado River below Glen Canyon Dam. For these people, total value would consist entirely of existence value and the foregoing formulation of a conditional existence value would be an appropriate measure of economic value. However, a different formulation is required when total value may be affected by impacts to use values from changes in the value of recreation and changes in the price of electricity and possibly other marketed products. Let p and Q be the price and resource quality vectors under current operating criteria and p′ and Q′ hold under the alternative criteria being evaluated. As before, we assume that, on net, the consumer will be better off, so that the compensating measure of willingness to pay is appropriate. Following Randall and Stoll (1983), we define total value, TV, implicitly from the following relationship:

$$V(p, I, Q) = V(p', I - TV, Q').$$

So long as we maintain the assumption of certainty, TV, as defined above, incorporates all the relevant effects: the value of the electricity price increase, the value of improvement in recreational quality, and the existence value. For an individual who is linked to the Colorado

River in only one way (i.e., as a power consumer or a recreationist or a non-user concerned about existence), TV collapses to one of the measures previously defined. If the Colorado River resources at issue affect well-being in two or more ways, then distinguishing between the components of total value is more complicated. For example, recreation and power consumption or recreation and non-use values could be related through relationships of substitutability or complementarity. Using a resource might make one's value of preserving it for future generations higher than if one were a non-user. In such cases, a problem of "jointness" makes dissecting TV into unique components conceptually impossible. For lack of a better way to do it, if separating existence value were necessary, the approach that goes back at least to McConnell (1983) must suffice. Letting UV equal use value,

$$UV = TV - EV_0$$

For the non-user, EV_0 equals TV and there is no problem. For the user, any jointness between use values and existence values is fully reflected in UV. At first, this may seem like an unimportant point. Who really cares about a consumer's motivations for holding UV? However, if one were to measure a user's value using the travel-cost method and contingent valuation, the two might not be equal, since the existence value generated by recreation might not be fully reflected in the number of trips. Perhaps Hanemann was getting at this sort of problem recently when he remarked that a more global concept of weak complementarity, rather than one that applies only to the choke price, is needed.

In the study plan sketched below, however, this concern does not appear to be severe. The primary emphasis will be on total value rather than the component parts. Some thought will need to be devoted to the relationships between total values as measured in a separate study and the recreational and power values that have already been estimated or soon will be. Our recreation values for the impacts of alternative flow regimes reflect only the short-term effects of flows on non-market values and thus may not be much influenced by longer run environmental effects, although some thought will need to be given to whether we can test this hypothesis. A substantial link between electricity prices, within the range we are likely to be talking about, and either recreation demand or existence value seems unlikely, although again possible empirical tests might be considered.

To summarize, at this level of abstraction, existence values have a sound basis in economic theory. Existence values are easily integrated with other values in an internally consistent way so long as total value can be used as the welfare measure. So far, what we have done seems to us to be fundamentally consistent with all the major theoretical treatments of existence value in the published (or soon to be published) literature, including the more descriptive treatments found in Krutilla (1967) and Krutilla and Fisher (1975), and more formal treatments by Randall and Stoll (1983), McConnell (1983), Madariaga and McConnell (1987), Smith (1987), Boyle and Bishop (1987), Bishop and Welsh (1990), Brown and Plummer (1990), and Freeman (forthcoming). If existence-related effects are present in a utility function, the conventional logic of welfare measurement easily accommodates this new dimension.

Of course, up to now we have been very abstract. When one digs a bit deeper into the literature, some differences among the contributions of various writers come to light and should be addressed here. We shall examine, in turn, alternative views (1) about where to draw the boundary between use and existence values; (2) about the motivations for existence values; and (3) about how existence values should be interpreted in the context of public decision making.

Boundaries Between Use and Existence Values

Thus far, we have not clearly distinguished between use and existence values. Two traditions exist in the literature. One, going back at least to Krutilla and Fisher (1975) and continuing through Smith (1987), makes the break on the basis of whether in situ contact with the resource is involved. " In the case of existence value, we conceived of individuals valuing an environment regardless of the fact that they feel certain they will never demand in situ the services it provides. . ." (Krutilla and Fisher, 1975, p. 124). The other tradition, going back at least to McConnell (1983), is based on the economic concept of weak complementarity. Use values are values that include in situ use but also include off site uses that are linked through weak complementarity to market purchases. For example, the first tradition would categorize values of reading about natural environments at the bottom of the Grand Canyon or enjoying photographs or other visual presentations about them as existence value, while the second would

categorize such values as part of use values. In the latter tradition, Boyle and Bishop (1987, p. 944) call such values "indirect use values."

While this debate is of some theoretical interest, we propose to sidestep it for purposes of this report by focusing attention primarily on total value as defined above. Total value will include indirect use, whichever side of the boundary it falls. Economic work under GCES does include in situ uses of Colorado River resources: whitewater boating in the Grand Canyon, fishing below the dam, and power generation at the dam. However, we will not attempt to separate indirect uses as a separate category of value when working in a total valuation framework.

While existence values are relatively easily incorporated into the formal treatment of consumer welfare to form total value, more needs to be said about them. Because they are not revealed in market and other behavior to the same degree as use values, substantial attention has been given in the literature to the possible motivations for holding them. Normally, economists rely heavily on observations of market behavior to identify what should be included in the utility function. Lacking the ability to observe market behavior stemming from existence value, motivations are important indicators of underlying preferences. After discussing motivations, we will turn to some important conceptual issues associated with interpretation of existence value.

Motivations for Existence Values

Economists have a deep-seated tradition of basing economic values on individuals' tastes and preferences. If a member of society behaves as if his or her economic welfare is affected by some variable, that is normally taken as sufficient evidence that his or her welfare is affected by that variable. Once the concept of total value is adopted and the theoretical possibility of existence value is admitted, behavior becomes less than a perfect indicator of welfare. Membership in environmental organizations is often taken as an indication of existence values, although as Freeman (forthcoming) has recently pointed out, there are mixed messages that make it nearly impossible to infer much about existence values from contributions and memberships. Ultimately, the economic researcher must base judgments about whether

existence values are real or not at least partly on what people say, rather than what they do. We economists do not feel very comfortable about taking what people say as economic evidence. Nevertheless, we feel compelled to ask not only whether people hold values that are not rooted in use, but also why.

In addition, the study of motivations can help identify and characterize in service flows that are important to people. As noted previously, it is hard to know a priori what to put in the utility function to express the contribution of the resource to welfare. It is probably acceptable at a theoretical level to put Q in the utility function as we did here. However, as Brookshire et al. (1986) pointed out, that does not tell us very much about what people are really valuing. For example, how should the humpback chubs be viewed? Should we simply include the population of chubs in the utility function? It is intuitively plausible that the loss of a large share of the population to a natural cause, say a storm event, would have one value for the loss and an equal population reduction due to a human-caused event such as a chemical spill would have another, possibly much higher value for the loss. It is important for policy to know as much as possible about what people are really valuing when they express existence values.

Motivations may also be important because, as Madariaga and McConnell (1987) have shown, motives can have implications for the interpretation of existence values in a benefit-cost framework. As we shall see below, altruism is often taken as one of the motivations for existence values. Madariaga and McConnell showed that values based on certain forms of altruism could easily be incorrectly interpreted.

Finally, writers on existence values have been forced to consider the motives underlying existence values by what we shall here term the "Pac-Man syndrome." Many of our colleagues have viewed human beings as "consumers" for so long that they are highly suspicious of any concept of value that is not linked directly to personal physical consumption. Humans are viewed as machines moving through the maze of life as rapidly as possible and voraciously consuming everything in their paths until death intercepts them. A good example of this mentality is to be found in Mendelsohn's (1984) doubts about existence values.

> . . . there is reason to suspect that existence value may not even exist. After all, why would people value something with which they have no contact and for which they cannot anticipate contact. What difference would it make if it was not there? How would they even know it was not there when it ceased to exist? Clearly, if a lot of us possessed substantial existence value, it would give a shyster a lot of room to maneuver as he promised to preserve things but never did. Could we rightfully complain? Perhaps we could insist on third party verification that the creature remained. Would we pay a lot to hear a "yes," or would we want to know more. Perhaps a film of the creature and an occasional book would do. But if this is all we want to know of the creature's existence, what would stop the shyster from making several such films and books and then destroying the creature. It appears that most people's notion of existence value is probably another form of use value, and probably should not be added to direct and secondary use value.
>
> To test for existence value, it is necessary to eliminate potential use from consideration. For example, how much would you pay a millionaire who owned his own island to preserve some small fish in the middle of his property if it was clear that public access would never be granted to the area. . . . Casual empirical evidence suggests that true existence value is zero. (Mendelsohn, 1984, p. 10)

Mendelsohn treats bequest value separately, pointing to a possible double-counting problem:

> . . . the present value of use is the discounted value of all future use of the resource. It is very difficult to tell in what way bequest value differs from the string of discounted future benefits of users. Bequest value appears to be future user value called by a different name. . . . If future use is properly incorporated into direct use measures, bequest value is redundant and should be ignored. (Mendelsohn, 1984, pp. 10-11)

Taking up Mendelsohn's challenge, let us ask why people might place a value on maintaining a resource even if they would not personally benefit through use. Altruism has played a key role in the conceptual literature on existence value (see, for example, McConnell, 1983, and Randall and Stoll, 1983), and rightly so in our opinion. In an earlier paper, Bishop and Heberlein (1984) suggested that existence value might stem from several kinds of motives. One is benevolence toward relatives and friends. Giving of gifts to friends and relatives is very common and would appear to stem partly from altruism. Why should such activities not extend to natural resources use opportunities? If Alpha would enjoy knowing that her neighbor, Beta, has the opportunity to watch birds in a certain marsh, both could benefit from marsh preservation. If Beta actually goes bird watching there, he receives a use benefit, but, contrary

to what Mendelsohn seems to be saying, the value would not end there. Alpha would also benefit personally, and counting only Beta's use value would miss this existence value that accrues to Alpha.

Bishop and Heberlein also noted that existence value could be motivated by sympathy for and empathy with people and animals, by environmental linkages, by feelings of environmental responsibility, and by bequest goals. They pointed out (p.10):

> Even if one does not plan to personally enjoy a resource or do so vicariously through friends and relatives, he or she may still feel sympathy for people adversely affected by environmental deterioration and want to help them. Particularly for living creatures, sympathy may extend beyond humans.

Those who have watched the animal rights and anti-hunting movements cannot help but be impressed by the intensity of feeling that some people exhibit in that context, and potential future use values could hardly explain their motives. Such concerns may partly motivate Randall and Stoll's (1983) so-called Q-altruism. Environmental linkages relate to the "you've-got-to-stop-'em-somewhere" attitudes. Environmental concerns are widespread, and environmental events at Location A, which a given individual does not use, may cause her/him to feel more or less confident about events at Location B, which the individual does use. Motives based on feelings of environmental responsibility have to do with people's concerns about the effects of their consumption on environments that they do not personally plan to use. For example, if Gamma's consumption of electricity would contribute to deterioration of Grand Canyon beaches, then she might be willing to pay something to reduce or eliminate this effect so that she is not responsible for such harm. Bequest motives are a temporal extension of motives relating to benevolence toward relatives and other people into the temporal realm. Again, it seems that Mendelsohn and others miss the point. Yes, the beneficiaries may well receive use benefits and those use benefits are quite correctly counted. The point, however, is that the benefits do not end there. If the benefactor's utility function depends on the bequest, an additional value is created and this additional value is missed if the beneficiary's use value alone is included in benefits.

Of course, the ultimate actuality of existence values and their underlying motivations are empirical issues. The point here is that they cannot be ruled out a priori. Nor do we see any merit in the direction taken in the previously mentioned paper by Brookshire et al. (1986). They tried to argue that existence values, though they could represent real willingness to pay, should not be considered as economically relevant because they may reflect ethical considerations other than the efficiency ethic that underlies benefit-cost analysis. To the contrary, ethical considerations must be viewed as very important foundations for the tastes and preferences that influence economic activity within the market and outside. Brookshire et al. try to establish that commitments to some ethical stands can lead to "counter-preferential choices" (p.1510 and elsewhere), but surely this runs contrary to the whole concept of revealed preference. No one would propose discarding a portion of the compensating variation associated with the availability of a market good at a stated price simply because some purchasers wish to use the good for charitable purposes. Why try to establish such an exception for the existence of public resources simply because those who are concerned about it are basing those concerns on altruistic motives? Similar objections should be raised with respect to Kahneman and Knetsch's (forthcoming) argument that existence values should be disregarded because they involve the purchase of "moral satisfaction."

The consideration of motivations does help us address one issue that has already been raised by some people in the context of GCES. Why do we limit existence values to the environment? Isn't it just as conceivable that non-environmental items might also have existence values? The case in point is electric power. It seems to us that if alternatives under consideration as part of GCES would involve increases in the cost of power or affect reliability of service or otherwise affect the final consumers of electricity, an existence effect cannot be ruled out a priori. One might speculate that the effects on the overall power system, in terms of the prices of power to final consumers, reliability of the system, etc. will be so small that they will be of little consequence for non-use values. However, if the power system modeling suggests that there will be large impacts to consumers of electric power and people care about these impacts for purely altruistic motives, consideration will need to be given to how this might be addressed empirically in a total valuation framework.

To summarize the theoretical discussion so far, there seems to be a growing consensus about the basic structure of the theory of existence values and how existence values are related

to other categories of values. Some minor differences in the taxonomy of values remain but they do not appear to be relevant to the study contemplated in this report. Because there are no market or market-related behaviors associated with existence values, the study of existence values has inevitably drawn economists into the unfamiliar area of motivations. Nevertheless, a fairly well-articulated set of hypotheses about the motivations underlying existence values has evolved. Nothing we have considered so far would rule out existence values as theoretically valid values that are fully integrable with other economic values within a total valuation framework. We now turn to a different set of issues that surround how existence values ought to be interpreted in the context of public decision making.

Interpreting Existence Values

The decision to include existence values as a valid portion of total value does raise some issues of interpretation. To explore these issues let us propose, as a working hypothesis, that one important source of existence values is a general concern about the environment. Polls have repeatedly shown that such concerns are widespread in the United States and have remained so for many years. If an individual, Delta, has a general concern about the environment, he might be willing to pay something to protect Grand Canyon riparian resources because this would be one way "to do something" about environmental degradation and its effects on wildlife, which he sees as deleterious to his well-being. Even before we begin to worry about what is being measured in contingent valuation studies, we can see here one aspect of existence values that troubles some economists. Assuming that Delta's expression of willingness to pay stems from this source, does it really represent the value that he places on the Grand Canyon resources or does it represent some broader value that he holds for preservation of the environment as a whole?

Perhaps this question can best be addressed on the theoretical level using the concept of substitutes. Economic values always depend on the prices and availability of substitute goods and services. The value one places on one's automobile depends on the prices of other automobiles and other means of transportation. The higher the prices of these substitutes, the higher value one would place on his or her car. This principle is certainly true for use values

associated with non-market goods such as the recreationally exploited trout population just below Glen Canyon Dam. As an example, Alpha's consumer surplus for guided fishing trips in Glen Canyon depends on the price of such trips and the prices of other substitute recreational activities available to Alpha.

Such effects may be present for existence values as well. Preservation of Grand Canyon riparian resources might be considered as only one of many ways to satisfy consumer demand for protection of environmental resources that are not used by the person under consideration. Improving water quality at location X, cleaning up the air at location Y, or protecting wildlife habitats at location Z could reasonably be viewed as substitutes for protecting riparian resources in the Grand Canyon. Just as in the cases involving market and use values, the strength of the effect of any one environmental protection project on the values of others will depend on the extent of the substitutability (or complementarity) between the alternatives.

If this is an adequate picture of how existence values for environmental protection work, then it does not present a new theoretical problem. The non-use value of the Grand Canyon resources in question will be more or less depending on what is happening with respect to environmental quality at locations X, Y, Z, and elsewhere. Just as the value of an automobile depends on the prices of other automobiles and other methods of transportation, so the value of avoiding damage to the riparian ecosystem below Glen Canyon Dam depends on the availability of substitutes that will satisfy a demand for environmental protection. Delta's demand for transportation, given the price and availability of substitute forms of transportation, is satisfied partly through the automobile he currently owns. Likewise, his demand for environmental protection, given the price and availability of substitute forms of environmental protection, could be partly satisfied by suitable modifications in Glen Canyon Dam operations that will reduce damages to the resources in question. The theory applies to both use and non-use values (such as existence value) in a parallel fashion.

Nevertheless, when the theory is applied, some issues do command specific attention. We will deal with them here as the "project-selection problem" and the "adding-up problem." If there are a great many ways to satisfy the demand for environmental protection, and the different ways of doing so are considered to be nearly perfect substitutes, then this would have important implications for benefit-cost analysis of projects to achieve alternative forms of

environmental protection. "Project" is used here in the very broadest sense to include alternative policies and regulations, as well as activities involving capital expenditures. Changing the criteria for operating Glen Canyon Dam would be considered a "project" for purposes of this discussion. The project-selection problem occurs because many different environmental protection projects would be capable of satisfying environmental demands, at least to some degree. That the benefits of cleaning up the water at location X exceed the costs may not be a very convincing economic case for doing so if benefits of cleaning up the air at Y and of preserving wildlife habitat at Z exceed the costs of projects at those locations. Within a budget constraint, the problem is not simply to identify a project having positive net benefits, but rather to find out which combination of projects satisfies environmental demands in the most economically efficient way. To the extent that projects at X, Y, and Z are nearly perfect substitutes, the project-selection problem boils down to one of finding the most cost-effective way of satisfying some part of the demand for environmental protection within limited public budgets.

The adding-up problem is closely related to the project-selection problem. Suppose that, in separate studies, we accurately estimate the existence values for water quality improvements at X, air quality improvements at Y, wildlife habitat protection at Z, and all the other environmental "good things" that might generate existence values. The sum could be a very large number, perhaps an implausibly large number.

However, once we recognize that the different environmental protection measures might be substitutes, we see that Delta's willingness to pay for the three projects combined would not equal the sum of the individual project values. It would be theoretically incorrect to add up the three values calculated in isolation. Theory would lead us to expect that Delta would be willing to pay less for the three projects combined than the sum of the three individual project values.

This problem is certainly not limited to existence values, but is inherent to valuation generally (Hoehn and Randall, 1989). The value a person would place on a rapid transit system in his city will be different depending on whether on not he actually owns an automobile. The economic logic of calculating values is the same regardless of whether market or non-market values are at issue and whether the values are use values or existence values.

Note also that the adding-up problem is not too severe when dealing with today's values. Today's values are dependent on the prices and availability of substitutes (and complements) as they exist today. In our example, the current value of improving water quality at X, depends on actual conditions at Y and Z. But when we try to make projections into the future, values then will depend on whether Y and/or Z have been added in the meantime. Who can say what sorts of environmental protection measures might be put in place in the future that would reduce the future value of X. Likewise, environmental problems that are unknown or non-existent today may enhance the value of X in the future.

Whether our working hypothesis of high substitutability for the existence of many environmental good things is valid is an empirical question that needs to be investigated. In the meantime, thinking about the possible implication of the project selection and adding-up problems for GCES, two points need to be made. First, as has already been stressed, this is not a problem that is unique to existence values. It is true to a greater or lesser extent for all efforts to establish economic values. Second, at least some of the effects that we are addressing in GCES involve potentially irreversible impacts on the relatively unique natural and cultural resources of Grand Canyon National Park. Concerns relating to adding-up and project selection become less severe as the uniqueness of the resources in question increases. Uniqueness reduces substitutability. Even writers on the topic who have been most skeptical of existence values have admitted the potential relevance of existence values in such cases. For example, in remarks specifically focused on non-use values associated with damages to natural resources from spills of oil and toxics, Harrison and Hausman (1989, p.10) asserted that,

> the site damaged should be, in some important sense, unique and well known or significant, if not at a national level, then at the regional level. . . . Similarly, it makes little sense to assign existence value to individual animals of a species if the losses involved pose no threat to the survival of the species in the general area.

Beers et al. (1989, p.15) conclude that, "there is wide agreement that the concept of non-use values for most practical purposes relates only to certain unique, irreplaceable resources." In support of this conclusion they cite an attached statement by Freeman who pointed out (p.3),

The long literature on nonuse and existence values emphasizes the following issues: the uniqueness or specialness of the resource in question and the irreversibility of the loss or injury. For example, economists have suggested important nonuse values in preserving the Grand Canyon in its natural state and in preventing the global or local extinction of a species or the destruction of a unique ecological community. On the other hand, resources such as ordinary streams or lakes or a subpopulation of a widely dispersed wildlife species are not likely to generate meaningful nonuse values, because of the availability of close substitutes.

We suspect, given our discussions of potential motivations, that limiting existence values to irreversible effects on unique resources is premature and difficult to justify without empirical evidence. However, in the current context, such concerns may be irrelevant. The point is that the more irreversible the effects and the more unique or special the resources involved, the more relevant are existence values.

All this has important implications for empirical work under GCES. Some individuals working for federal agencies have suggested that when people talk of the uniqueness of Grand Canyon National Park, they have in mind the dramatic views and geological wonders visible from the rims and along the trails of the park. They would argue that most people are not particularly interested in the riparian resources in the bottom of the canyon and would not include those resources when speaking of the unique resources of the Grand Canyon. Others would argue that riparian resources are viewed by many members of the public as having existence values because they are interested in preserving ecosystems of the park as well as the visually and geologically unique resources. Stated differently, significant numbers of people may feel that the resources that are the focus of GCES take on "specialness" or uniqueness simply because they are part of Grand Canyon National Park in addition to whatever value they might have if they were not part of that broader set of resources. The role of Colorado River resources in people's perceptions of the park therefore needs to be investigated as a part of any research into existence values. Results will be relevant for evaluating the potential severity of the adding-up and project selection problems in interpreting the economic value arising from any total valuation study conducted under GCES.

Up to now, the discussion has assumed certainty. As a final step toward total valuation, in the next section we consider option value by modifying the valuation model to allow for uncertainty.

Option Value

Weisbrod's original paper on option value was so influential because its point was so intuitively appealing. Using Sequoia National Park as an example, Weisbrod posed the problem of whether the trees in the park should be cut if the value of the resulting wood products exceeds the value that consumers place on use of the living trees for recreational and aesthetic enjoyment. He reasoned that there could be many people who are not currently using the trees for recreation and aesthetic enjoyment who would nevertheless place a value on the option of using them for such purposes in the future. However, options to enjoy Sequoia National Park in the future cannot be purchased in the market. Once cut, the trees cannot be replaced for centuries if demand for their enjoyment turns out to be large in the future. Maintaining the option of future enjoyment takes on public goods characteristic in that not cutting the trees makes the option of future enjoyment available to all. Hence, economic assessment of the relative values of cutting the trees and preserving them would not be complete without including, on the preservation side of the ledger, the value of the options for future enjoyment.

For such a seemingly obvious point, Weisbrod's conclusion has generated a large literature. More than a dozen important contributions are easy to identify (Long, 1967; Lindsay, 1969; Byerlee, 1971: Cicchetti and Freeman, 1971; Schmalensee, 1972; Graham, 1981; Bishop, 1982; Freeman 1984, Chavas, et al., 1986; Smith 1987b; Hartman and Plummer, 1987; Cory and Saliba, 1987) and many other papers could be cited as well. The issue, and it has turned out to be a difficult one, has been the relationship between willingness to pay for options and the traditional welfare measures as discussed earlier in this report. At its root, the debate has centered around the appropriate way to measure economic values of individuals who are uncertain about future values of economic parameters affecting their welfare. Some basic definitions and theoretical conclusions have emerged.

First, let us revise the indirect utility function to be V[P(s), I(s), Q(s), s]. The variable s is Added to represent various states of the world that may occur in the future. Prices, income, the status of service flows from the resource, and preferences themselves may depend on the state of the world. Thus, V(·) is a conditional indirect utility function. For each state of the world, s, we can define a compensating measure of the welfare change, C(s), in the usual way,

$$V\,[P(s), I(s), Q(s), s] \;=\; V\,[P'(s), I'(s) - C(s), Q'(s), s],$$

where P(s), I(s), and Q(s) represent the values of economic parameters in the absence of the modification in dam operations or other steps to reduce adverse impacts of dam operations. Adding primes to these symbols signifies the values of the parameters if the modifications were made or other steps taken. Letting E symbolize the expected value operator over s, E[C(s)] is the expected value of C(s). The Hicksian equivalent measure of value could be defined in a parallel fashion. Once uncertainty is introduced, such values are termed ex post values because they are realized only after the state of the world, s, is known. In the terminology being used here, these are ex post measures of total value.

An alternative measure of the welfare to the ex post measures C(s) (one value for each state of the world) and E[C(s)], is option price (OP), which measures the total value ex ante, before the state of the world is known. To simplify the exposition, assume that the consumer maximizes the expected value of utility. If

$$E\{V\,[P(s), I(s), Q(s), s]\} < E\{V\,[P'(s), I'(s), Q'(s), s]\}$$

then the consumer is said to be better off ex ante if the proposal is adopted. If the inequality is reversed, he or she is said to be worse off ex ante. In either case, the compensating measure of OP is implicitly defined as:

$$E\{V\,[P(s), I(s), R(s), s]\} \;=\; E\{V\,[P'(s), I'(s) - OP, R'(s), s]\}$$

The option price is a state independent payment. That is to say, OP does not depend on s. If the proposal would increase the expected value of utility, OP is interpreted as the maximum sure payment that this consumer would pay ex ante (before the state of the world is

known) to see the alternative in question adopted. If the proposal would reduce the expected value of utility, then OP would be negative and is interpreted as the minimum sure compensation the consumer would have to be paid ex ante to get him or her to acquiesce to the adoption of the proposal.

Finally, OV, which symbolizes option value, has come to be defined as

$$OV = OP - E[C(s)]$$

That is, option value is taken to be the difference between option price (the ex ante measure of total value) and the expected value of the ex post measures. It might be interpreted roughly as an adjustment for the risk preferences of the consumer. Risk preferences affect how the consumer assesses the relative risks of paying the option price ex ante and signing a contract to pay C(s) ex post, where the value of C(s) will depend on the particular state of the world that ultimately occurs. It is now well known that the sign of option value cannot be predicted a priori except in special cases. Furthermore, it is conceivable that option value could be substantial in absolute value (Freeman, 1984). Many today are questioning whether expected utility maximization is an empirically justifiable assumption. If expected utility maximization is rejected and a more general definition of option price adopted, then option value could conceivably take on any value, positive or negative, large or small, depending on how consumers actually deal with uncertainty.

It is now clear that option value is not a separate value at all, but the difference between two alternative welfare measures under uncertainty. Defining the non-use portion of total value as option values plus existence values is not justified.

Just how welfare changes ought to be valued in the presence of uncertainty is not a simple question. One approach would be to use ex post measures such C(s) or E[C(S)]. Alternatively, some ex ante measure like option price could be used. The trend in the literature seems to be in the direction of ex ante measures (Graham, 1981; Cory and Saliba, 1987; Bishop, 1986; Freeman, 1986; Smith, 1987b; Freeman, forthcoming). The problem is that ex ante measures other than option price have been suggested. Without going into detail, let us simply suggest that this work is not sufficiently mature to draw ironclad conclusions. In our judgment, a

a concept like option price that allows for additional theoretical issues such as project financing will ultimately win out (Ready, 1990). In addressing the empirical problems associated with GCES, we would propose to let option price serve as the guiding theoretical construct.

The most important implication of all this for GCES has to do with the wording of contingent valuation questions. To the extent that it is possible to do so, contingent valuation scenarios should be written from an ex ante perspective so that they will reflect option prices rather than ex post values.

Taking Stock

Economic theory is very important for applied economic analysis. It guides the questions asked, the analyses performed, and the interpretation of results. If we had found that the theory of non-use values was poorly developed or poorly integrated with the main body of accepted theory or that economists were engaged in a full scale theoretical debate over fundamental issues, then this would have posed major questions about the wisdom of conducting non-use value studies under GCES. Instead, non-use values have been easily integrated into the body of welfare theory through the concept of total value. To the extent that there is any theoretical debate, it tends to focus on taxonomic issues regarding the boundary between use and non-use values and the possible motivations for existence values. We have been unable to locate a single attack on the theoretical fundamentals of existence values that is having or should have much influence on the thinking of those who are leading in the conceptual development of existence values. Where we find potentially relevant theoretical issues in the interpretation of existence values for public decision making, as in the case of the adding-up problem, it becomes apparent that such issues are not unique to existence values but are present to a greater or lesser extent for market and non-market use values as well.

Matters are somewhat less settled once uncertainty is introduced. The process of convergence toward widely accepted welfare measures under uncertainty is still evolving. Nevertheless, considerable progress has been made toward agreement that ex ante measures are appropriate. Further refinements to incorporate such issues as project financing are in the

offing. In the meantime, option price will serve as a sound theoretical foundation for current applications.

At first glance, this rather optimistic view seems at odds with statements that one finds occasionally in the literature. For example, Freeman (forthcoming, p.2) has stated, ". . . there is very little agreement among economists as to terminology, definitions, [or] what motivates people to hold nonuse values. . ." Such conclusions are at least partly motivated, not by the literature on which we have built Section 1 of this report, but by the more empirically oriented literature that will form the basis for Section 2, below. Substantial progress has been made on the theory of total valuation during the last decade. Some of the empirical studies of non-use values were done before the theory had evolved to its current point. In other cases, applied studies have been performed without careful attention to theoretical concerns. The result is flawed research.

Our conclusion is that the theory is sufficiently well developed and integrated with the main body of economics to form the foundation for a total valuation study under GCES.

While accepting the theoretical validity of non-use values, it is important to realize that they can be included in real world evaluations only if they can be measured reliably. Because of their very nature, existence values can not be measured using any technique relying on weak complementarity. In the present state of resource and environmental economics, this means that measurement of total values will require the use of the contingent valuation method. However, the ability of contingent valuation to reliably measure economic values has been the focus of substantial debate. The next section will review the contingent valuation literature with specific attention to studies exploring the accuracy of the method and past applications of the method to the problem of measuring total values.

SECTION 2: CONTINGENT VALUATION OF TOTAL VALUES

As discussed in the theory section, total value is comprised of use and non-use values. Depending on the perspective used to define existence or non-use values all (or at least a portion) of non-use values can not be measured using direct market based methods nor any of the non-market valuation methods based on the assumption of weak complementarity. As a result, empirical studies of total value in which non-use values comprise some or all of the total value, have relied on the contingent valuation method. In the first part of this section we will review studies that have provided estimates of non-use values. It is important to bear in mind when reading this section that some of the studies reviewed were conducted rather early in the development of the contingent valuation method and the theory of non-use values. As a result, while these studies will be illustrative of the magnitude of non-use values found in various settings, they do not all adhere strictly to the definitions used in this report.

There has been substantial debate among economists (and other social scientists) over the prospects that the hypothetical expressions of willingness to pay collected using the contingent valuation method provide valid inferences about the economic value of the non-market good being studied. This debate has sparked several attempts to compare the values obtained using contingent valuation to values obtained in experimental markets in which actual transactions occur. Because of the central role that contingent valuation plays in any study of non-use values the validity of the contingent valuation method is a key issue. To this end, the second part of this section contains an overview of literature addressing the validity of the contingent valuation method.

This section will close with a discussion of the issues raised by the literature review, which must be addressed in any study of non-use values.

Table 1: Review of Studies Presenting Non-Use and Use Value Estimates

Source	Item Valued	Non-Use Value	Use Value	Ratio of Non-Use to Use Value
Meyer (1974)[a]	Preservation of salmon, Fraser River, British Columbia	$502.00	$928.00	0.54
Horvath (1974)[a]	Wildlife attributes, Southeastern U.S.	$1,574.00	$2,824.00	0.56
Dornbusch & Falcke (1974)[a]	Water quality improvement, U.S. water bodies	$--	$--	.75-2.03
Meyer (1978)[a]	Preservation of salmon, Fraser River, British Columbia	$360.00	$287.00	1.26
Walsh et al. (1978)[a]	Water quality improvement, South Platte River, CO	$66.00	$126.00	.53
Mitchell & Carson (1981)[a]	National water quality improvements	$121.00	$258.00	.47
Desvousges, Smith, and McGivney (1983)[a]	Changes in water quality, Monongahela River	$34.00 $28.00	$52.00 $62.00	.65 .44
Cronin (1982)[a]	Changes in water quality, Potomac River	$30.00 $84.00	$42.00 $125.00	.71 .67
Cronin (forthcoming)[a]	Changes in water quality, Potomac River	$35.00 $66.00	$44.00 $137.00	.80 .48
Brookshire et al. (1983)	Increase in Grizzly population in Wyoming requiring 15 years	$15.20	$5.80	2.62
	Increase in Grizzly population in Wyoming requiring 5 years	$7.40	$15.60	0.47
	Increase in Grizzly population in Wyoming requiring 15 years	$6.90	$11.10	0.62

Table 1: Review of Studies Presenting Non-Use and Use Value Estimates (Continued)

Source	Item Valued	Non-Use Value	Use Value	Ratio of Non-Use to Use Value
Walsh, et al. (1984)	1.2 million acres in wilderness areas in Colorado	$10.9[f]	$17.6[f]	0.62
	2.6 million acres in wilderness areas in Colorado	$14.6[f]	$27.0[f]	0.54
	5 million acres in wilderness areas in Colorado	$19.7[f]	$41.2[f]	0.48
	10 million acres in wilderness areas in Colorado	$24.8[f]	$68.4[f]	0.36
Walsh, et al. (1985)	Presentation of 11 wild and scenic rivers in Colorado	$61.00	$33.00	1.85
Sutherland and Walsh (1985)	Presentation of present water quality levels in Flathead, MT	$46.25	$18.08	2.55
Hageman (1985)	Whales	$22.94	$2.34	9.80
	Dolphins	$16.35	$2.21	7.40
	Otters	$18.33	$2.49	7.36
	Sea Elephants	$17.66	$1.16	15.22
Stoll and Johnson (1985)	Whooping Crane Survival	$1.03	$1.40	0.74
Boyle and Bishop (1987)	Bald Eagle presentation in Wisconsin	$28.00	$47.00	0.60
King, et al. (1988)	Desert Bighorn Sheep population near Tuscon, AZ	$24.00	$-2.20	0[e]

Table 1: Review of Studies Presenting Non-Use and Use Value Estimates (Continued)

Source	Item Valued	Non-Use Value	Use Value	Ratio of Non-Use to Use Value
Loomis (1989)	Mono Lake, CA: Alternative 1 versus Alternative 2[c]	$4.72[d]	$6.70[d]	0.70
	Mono Lake, CA: Alternative 2 versus Alternative 3[c]	$4.12	$8.03	0.51
Boyce, et al. (1989)	Norfolk Island Pines	$2.00	$4.81	0.42
Welle (1990)	Presentation of Minnesota Lakes from acid rain damage	$65.00	$67.00	0.98

[a] See Fisher and Raucher (1984).

[b] This ratio represents the simple mean of 11 different ratios from the eight studies as presented by Fisher and Raucher (1984) Tables 1 and 5. For one of the studies, Fisher and Raucher presented a range of ratios and we used the mid-point.

[c] See text for explanation of the alternatives.

[d] Loomis (1989) obtained values from both an "original" and a "retest" survey to test the reliability of contingent valuation. We present original values here.

[e] Since a negative use value would be impossible and since the difference between the existence value of non-observers ($21.80 with standard error estimated at 2.6) and the total value of observers ($24.00 with estimated standard error or 5.0) are not statistically different, we set this ratio equal to zero rather than -10.90.

[f] Unlike elsewhere in this table, where values are presented on a per household or individual basis, these figures are measured in millions of dollars for all Colorado residents combined.

Brookshire, Eubanks, and Randall (1983) estimated option prices and existence values for grizzly bears and bighorn sheep in Wyoming. Values were collected using open-ended contingent valuation questions. This study was based on a sample of Wyoming residents holding elk, deer, or antelope hunting licenses. In the questionnaire respondents were asked if they ever planned to hunt or observe the species in question at any point in the future. The values elicited from individuals with no plans to hunt or observe the species in the future were interpreted as existence values. Values solicited from all others were interpreted as total values that could potentially include current use, certain future use, uncertain future use, and existence values. Values in the table are for portions of the sample that claimed they would not hunt grizzlies and bighorn sheep even if they had an opportunity to do so. Values involving improvements in the populations requiring five and fifteen years are given. "Observer" values apply to individuals who expressed the expectation of observing grizzlies and/or sheep if the population were improved. As before, we assume that the existence values of observers and non-observers are equal. To see how the ratios were figured, consider the Grizzly bear numbers for a fifteen year time horizon. Mean annual existence value from survey results for non-observers was $15.20. Total value for observers averaged $21.00, implying, under the assumption that existence values of observers equals that for non-observers, that use value must be $5.80. Thus, the ratio is $15.20/$5.80 or 2.62. The negative use value for grizzlies with a five year time horizon is either a statistical fluke or raises questions about the assumption that observer existence values equal non-observer existence values. If, as this result would imply, non-observers have higher existence values for these species, then the ratios are overestimated across the board for the Brookshire et al. study.

Next we turn to a series of studies performed by Richard G. Walsh and various students and colleagues. This series of studies all followed very similar practices in applying contingent valuation. Respondents were first asked using an open-ended contingent valuation question, to state their total willingness to pay for the resource under study. Respondents were then asked to allocate their total value among use values, option values, existence values, and bequest values in percentage terms. This practice raises certain theoretical and survey methodological misgivings, but these do not rule out their usefulness in illustrating the range of empirical estimates of non-use values. One problem that needs to be considered in order to ensure greater consistency with the other results in Table 2 is the definition of option value used in these studies. Respondents were asked to allocate a portion of their total values to reflect their

willingness to pay for future use of the resource. Thus, responses to this part of the exercise should be interpreted as option prices rather than option values under the more standard theoretical definitions used in this report. Such option prices are likely to contain some future use values and, indeed, may be dominated by future use values, with option value proper perhaps playing a small role. Thus, in calculating the ratios in Table 2, we lump current use values and option prices ("option values" in the terms of Walsh and his co-authors) in the denominator. The ratios are then to be interpreted as the existence value under our definition (including both existence and bequest values of Walsh and co-authors) divided by use value, which includes current use values plus respondents' option prices for future use.

Walsh, Loomis, and Gillman (1984) estimated preservation values (including "option values" for future use and existence values, where the bequest values were estimated separately) for the preservation of wilderness areas in Colorado, using contingent valuation. The study was based on a random sample of Colorado households. In the questionnaire, respondents were asked to state a value that they were willing to pay for preservation of four different total acreages of wilderness in Colorado. This particular study deviated somewhat from the procedures in the other Walsh studies in that current use values were estimated using a travel-cost model. The travel-cost method provided use value estimates on a per activity day basis while contingent preservation values were on a per household per year basis. Walsh et al. calculate the aggregate use benefits for wilderness and the aggregate non-use benefits of wilderness to Colorado households and these values are presented in the table and are used to calculate the four ratios presented in the table. The contingent valuation portion of this study was based on a mail survey sent to a random sample of 218 households. The response rate was 89 percent.

Walsh, Sanders, and Loomis (1985) focused on the preservation of wild and scenic rivers in Colorado. Here, current use values and preservation values (option price and existence and bequest values) were all estimated using the contingent total values and percentage allocations as described above. This study was based on a questionnaire mailed to 214 randomly selected Colorado residents.

Sutherland and Walsh (1985) estimated the willingness to pay of a sample of residents of four Montana cities for maintaining currently high levels of water quality in Flathead Lake,

Montana. Total values were elicited and divided into various use and non-use categories using percentages. This study employed a mail survey with a sample size of 280. The response rate was 61 percent.

Hageman (1985) studied the value of marine mammals in the coastal waters of California. The study employed a payment card version of the contingent valuation method. Outliers were identified and eliminated from the analysis as were individuals specifying a zero willingness to pay. Respondents were asked to allocate their willingness to pay among three categories, non-consumptive uses (viewing), maintaining options for potential future use, and preservation of the species without any possibility of viewing the animals. For the purposes of this review, the amounts allocated to the last two categories were classified as non-use values. The ratio of non-use to use values ranges from 15.22 per household for sea elephants to approximately 7.40 per household for dolphins and otters. The Hageman study was based on a mail survey sent to a random sample of 1,000 California households. Hageman reported an overall response rate of 21 percent.

Stoll and Johnson (1985) examined the value of preserving whooping cranes. The study employed a sample consisting of three groups: visitors to Arkansas National Wildlife Area in Texas, which is the wintering grounds of the whooping crane, a sample of Texas residents, and a sample of residents of four major U.S. cities outside of Texas. Using the dichotomous choice technique of the contingent valuation method, Stoll and Johnson estimated use and non-use values and presented several different estimates depending on the type of value and the group. The data were further analyzed in a paper by Bowker and Stoll (1988), but the Stoll and Johnson figures are better suited to the sort of comparisons we need to make here. The figures in Table 2 were derived as follows. The first valuation question addressed to Arkansas visitors asked about willingness to pay for an annual permit to visit Arkansas. Responses were analyzed to estimate an average annual use value of $4.47 per visitor. Next, Arkansas visitors were asked about their willingness to pay for this same permit if there were no opportunities to view the cranes during visits. The average value here was $3.07, implying that the cranes contribute about $1.40 per year to use values of the refuge. Another question addressed to all three groups surveyed asked respondents to assume that government programs to maintain the whooping crane population were to be terminated, "a decision which would virtually insure the extinction of the whooping crane" (p.389). Respondents were then asked about their willingness

to pay membership fees to "an independent foundation that would purchase and maintain refuge land so that the species might be preserved for the future" (p.389). The responses of greatest interest to us here are those of Texas residents who did not expect ever to visit Arkansas (averaging $1.03), since a pure existence value for whooping cranes would be the most plausible interpretation of the results based on their responses. The ratio given in the table should be interpreted as the ratio of the pure existence value of Texas non-users of the whooping crane to the use value of current users. However, it is interesting to note that visitors to Arkansas reported an average existence value of $9.33 while the residents of the four out-of-state cities reported an average existence value of $1.24. This may constitute evidence that individuals having direct personal experience with the resource in question tend to have higher non-use values. The Stoll and Johnson research was based on initial samples of 800 surveys distributed to visitors at Arkansas, 1,200 to Texas residents, and 600 to non-Texas metropolitan area residents. Stoll and Johnson report a response rate of 67 percent for Arkansas visitors and a response rate of 36 percent overall for the 1,800 mail surveys sent to non-visitors.

Boyle and Bishop (1987) estimated the value of preserving two Wisconsin endangered species: the bald eagle and the striped shiner. For the purpose of understanding the relative magnitude of use and non-use values, the values associated with preservation of bald eagles are most relevant. They estimated the total value of preserving bald eagles as well as a conditional value of preserving bald eagles given that the preservation would occur only in areas offering no opportunities for viewing the eagles. A random sample taken from Wisconsin tax roles was split according to whether or not the individual had contributed to a wildlife preservation fund and whether or not the individual had ever taken a trip whose purpose had been to observe bald eagles. Values were estimated using the dichotomous choice format of the contingent valuation method. The total value figures for those who had taken trips to view eagles were interpreted as a combination of use values, option values, and existence values. The conditional values for both groups can be interpreted as pure existence values. For those who had contributed to the wildlife preservation fund and had previously taken a trip to view eagles, the total value was $75 if eagles were viewable and $28 if the bald eagle were preserved in Wisconsin but could not be viewed. The $47 difference between these two values should be interpreted as use value plus possible option value for future viewing opportunities. The $28 can be taken as a measure of average existence value for the bald eagles in Wisconsin. The Boyle and Bishop study was based

on a mail survey of 1,000 Wisconsin residents. They reported an overall response rate of 81 percent.

King et al. (1988) focused on the value of preserving a herd of desert bighorn sheep in a wilderness area near Tucson. This study employed open-ended contingent valuation questions. A random sample of households in the Tucson metropolitan area was used to estimate the use and non-use value of the herd. The authors noted that very few of the current users of the wildlife area view bighorns during their trips, and that the viewing of the sheep did not seem to be an important motivation for trips. They suspected that existence value would comprise a significant proportion of the total value that individuals hold for the preservation of the sheep. The respondents were asked about their willingness to pay to preserve the desert bighorn sheep under two conditions. The first condition was that the sheep population would be maintained, rather than being lost due to human activities, and could still be viewed with about the same frequency as currently. The second condition was that the sheep population would be maintained, but people would be excluded from areas in which the sheep live. Values solicited under the first condition, which averaged $17.14 per respondent per year, were interpreted as total values of preservation and the values solicited under the second condition, averaging $15.14, were interpreted as existence values. This would imply use plus option values of $2.00 and a ratio of existence to use plus option value of 7.57, a very high ratio, but not a surprising one given how rare it is for the sheep to actually be seen. The King et al. study employed a mail survey. They report an initial sample size of 1,000, with a return of 550 surveys representing a response rate of 59 percent of deliverable surveys.

Loomis (1989) examined the use and non-use values associated with maintaining various water levels at Mono Lake in California. Three alternative lake levels, depending on water diversions, were valued. Alternative 1 involved low diversions and resulting high, biologically productive lake levels; Alternative 2 dealt with moderate ecological damage to the lake through moderate diversions and intermediate lake levels; and Alternative 3 involved high diversions, low lake levels, and resulting large environmental damages. Those surveyed in this study consisted of a sample of visitors to Mono Lake and a general population sample of California households. Values were estimated using the open-ended technique of the contingent valuation method. Values associated with California households were interpreted as consisting primarily of non-use values while the values reported for visitors reflect both non-use values and use values. Values

reported in the table assume that general population values are pure existence values, since relatively few Californians visit Mono Lake, and that Mono Lake visitors have the same existence values as Californians in general. The Loomis study employed two mail surveys. For the random sample of California households, Loomis reported a response rate of 44 percent for the 502 deliverable surveys. For the random sample of Mono Lake users, Loomis reports that 84 percent of 152 surveys were returned.

Boyce et al. (1989) studied the use and non-use value of Norfolk Island pine trees, a common houseplant. Because this study will be important in evaluating the validity of the contingent valuation method we review the basic features of their research plan in greater detail than for the other studies reviewed in this section. Subjects for this study were 115 staff members of the University of Colorado. Participants in the study were divided into four groups. Members of two of the groups participated in a contingent valuation exercise in which they were given a Norfolk Island pine and then asked the amount of compensation they would require to give it up. Members of the other two groups were asked to state what they would be willing to pay to acquire one of the pine trees. All participants were asked to state their bids as if they were participating in an auction. The rules of the auction were designed to be incentive compatible.

Existence values entered through a threat to destroy the trees not taken by the respondents. That is to say, participants in one of the willingness-to-pay groups were told that any trees not purchased by the study participants would be destroyed. Likewise, participants in one of the willingness-to-accept-compensation groups were told that trees they sold back to the researchers would be destroyed. The difference between values obtained when nothing was said about the trees and values obtained when the trees were threatened with destruction can be taken as a measure of the non-use values of the trees. Boyce et al. found that the threat to kill the trees increased the average willingness to pay from $6.06 to $16.80 and increased the average compensation demanded from $14.12 to $26.07.

As is often the case, the researchers encountered some difficulties in obtaining the compensation-demanded measures of the value of the trees. In particular, approximately 13 percent (four of thirty) of participants not getting information about the ultimate fate of the trees refused to state a bid in the hypothetical auction. When the trees were threatened with

destruction, nearly 50 percent (14 of 29) of participants refused to state a bid and one participant stated a bid of $5,000. Regardless of these difficulties, the researchers demonstrated that the ultimate fate of the pine trees had a significant impact on the value of the trees as collected using a contingent valuation exercise.

Welle (1990) studied the damages associated with acid deposition in Minnesota. The study employed both the dichotomous-choice and the open-ended contingent valuation techniques. The results were based on a random sample of Minnesota residents. The groups were defined in terms of the individuals' perception of the likelihood that they would engage in recreational activities in the area identified as potentially affected by acid deposition. Several different values could have been used in our effort here, but the results would have been quite similar. If acid rain damages to Minnesota lakes, as described in the survey, were certain to occur unless remedial steps were taken and if taking those steps were certain to prevent all damage, then respondents who indicated that they were certain they would use the undamaged lakes (referred to below as "certain users") indicated a mean willingness to pay to avoid the damage of $132. This would be interpreted as their total value, including both use value and existence value[1]. On the other hand, members of the sample who indicated that they were certain they would not use the lakes even if preservation were achieved ("certain non-users") indicated an annual value of $65, which would be interpreted as an existence value. If we assume that certain users have existence value equal to certain non-users, then that would leave use values of $67 ($132 minus $65). The Welle study employed a mail survey. Based on a sample of 910 deliverable surveys, Welle reported that 689, or approximately 76 percent, were returned.

Like the studies reviewed by Fisher and Raucher, the additional and highly diverse studies reviewed in this report consistently suggest that non-use values are positive and non-trivial components of total value. While the definitions are not altogether consist across these studies, both existence (including bequest) motives and motives relating to options for future use seem to be operative in supporting these expressions of value. The previous studies were selected because they all provided estimates of both use and non-use values for a wide variety of non-market goods. Given the theoretical difficulties identified in the previous section of

---

[1] There would be no option value here because there is no uncertainty

distinguishing between use and non-use values when both are present, these studies provide an important basis for understanding the range of magnitudes of the non-use component of total values when both use and non-use values are present.

Some of the studies just reviewed also provide estimates of non-use values for which there are no corresponding measures of use values, and there are a few studies in which only non-use values were measured. These additional studies are relevant to the GCES research because of the possibility that a substantial portion of non-use values may arise from individuals never planning to actually visit the area in which the environmental impacts of dam operations are felt. Studies providing estimates of non-use values are summarized next.

Schulze et al. (1983) studied the value of various levels of visibility at the Grand Canyon and for the southwestern region of the United States as a whole. They used an open-ended version of the contingent valuation method to elicit values for various levels of visibility. They asked two types of questions; one asked the respondent to state maximum willingness to pay for admission to the park (over and above the then current fee of $2) if visibility at the park was maintained at various levels. The second kind of valuation question asked about respondents' willingness to pay additional amounts on their electricity bill to maintain various levels of visibility. Average willingness to pay increased admission fees ranged from $5.38 to $8.79 per day. Average willingness to pay higher electricity bills to prevent reductions in visibility ranged from $3.72 to $5.14 per household per month. Schulze et al. do not provide sufficient information to provide separate estimates of use and non-use values. However, they note that the aggregate use benefits measured by willingness to pay increased admission fees are likely to be very small relative to the aggregate total benefits as measured by willingness to pay increased electric bills. They estimated the annual value of maintaining air quality at the Grand Canyon at approximately $6 billion per year. Furthermore, they state (page 173),

> The principal benefits of preserving visibility in the Grand Canyon Region. . . derive from the apparent desire of Americans to preserve a national treasure, whether or not they intend to visit or use the region themselves. Economists have turned this type of value "existence value."

As the next section will show, some may argue with the presumption that what was measured in this study was primarily existence values or even whether the value measured reflected visibility rather than the value of the Grand Canyon. However, the result reported in Schulze et al. demonstrates the importance of the population over which non-use values are aggregated. Even a small value per household becomes a very large number when aggregated over the total number of households in the United States.

In addition to reporting the value of the bald eagle, Boyle and Bishop (1987) also conducted a study of the value of the striped shiner. The striped shiner, a small minnow living in the Milwaukee River, is listed as an endangered species by the Wisconsin Department of Natural Resources. Because it is illegal to fish for these minnows and because of their habitat it is unlikely that any resident of Wisconsin will ever see a striped shiner. Using the dichotomous choice technique of the contingent valuation method, Boyle and Bishop estimated that the striped shiner had a value of $5.66 for contributors to Wisconsin's non-game fund and $4.16 for non-contributors. Both of these values can be interpreted as existence values.

Whitehead and Blomquist (1990) used a dichotomous choice version of the contingent valuation method to estimate the total value of preserving a specific wetland in Kentucky. Whitehead and Blomquist attribute the total value solicited primarily to existence value because only 15 percent of their sample had ever used a wetland for recreational activities. The purpose of their study was to determine whether specifically mentioning substitute wetlands had a significant effect on the values expressed by the respondents. They found that the value of preserving the wetland ranged from a low of $ .4 per household to a high of $39.99 depending on the description of the impacts to the wetland if preservation were not undertaken and whether or not other protected wetlands were mentioned as part of the valuation question. The authors conclude that the pattern of values observed during their study supports the notion that a description of potential substitute environmental goods is an important feature of any study of non-use values. The Whitehead and Blomquist study employed a mail survey of 1,000 Kentucky households. They report a response rate of 31 percent deliverable surveys.

Validity of the Contingent Valuation Method

The studies just reviewed give a strong indication that people will express a positive and sometimes large willingness to pay for an item they never plan to use either directly or indirectly. However, the question remains whether the hypothetical expressions of willingness to pay collected using the contingent valuation method are valid estimates of the value of non-market goods.

Economists, from the beginning, have been skeptical about the possibility of using surveys to value environmental resources and other goods and services. This skepticism is still very much alive today. Speaking about the potential for using contingent valuation in natural resource damage assessment, Phillips and Zeckhauser (1989, p. 520) claim, "CVM is unlikely to prove very accurate. This method is primarily useful for getting a rough cut at some otherwise unmeasurable value." Based on the alleged inaccuracy of the method, Phillips and Zeckhauser argue that contingent valuation should not be used to measure damages in environmental damage assessments.

In this section, we shall consider in more detail the accuracy of contingent valuation method and the implications of what we find for measuring damages to the ecosystem below Glen Canyon Dam as a result of dam operations. To focus the discussion more specifically on non-use values, two potentially damning arguments against the use of contingent valuation in the context of non-use values will be developed. First, during the mid-1980's, a panel of economists was commissioned to assess the state of the art in contingent valuation. Out of that effort came the so-called Reference Operation Conditions, or ROC's for short (Cummings et al., 1986). The ROC's claim to summarize the past research on contingent valuation in simple criteria describing when contingent valuation will work well and when it will not. The ROC's are of special concern here, since, as we shall show below, if they are accepted, they would rule out most applications of contingent valuation to non-use values. Second, in a paper forthcoming in the Journal of Environmental Economics and Management, Daniel Kahneman, a psychologist, and Jack Knetsch, an economist, describe evidence of what they term "embedding" effects and other phenomena in responses to contingent valuation surveys that claim invalidate the use of contingent valuation in most applications, including the estimation of existence values. This is a

particularly important challenge to contingent valuation because it represents an emerging set of concerns based not on economics, but on cognitive psychology. Having attempted to capture the essence of these two important lines of attack on contingent valuation, we shall evaluate them in light of empirical evidence on the validity of contingent valuation. Though contingent valuation continues to be controversial, there is, in fact, a growing body of evidence that supports the practical usefulness of resulting value estimates. Contingent valuation can be applied with some confidence to estimate use values, and early work on non-use values seems to be encouraging.

The Reference Operating Conditions (ROC's)

In the mid-1980's, Cummings, Brookshire, and Schulze (1986) undertook a comprehensive evaluation of the state of the art in contingent valuation. Their overview is particularly relevant given that all three, especially Brookshire and Schulze, have been pioneers in the development of the technique. They summarize their conclusions in the form of four "Reference Operating Conditions" or "ROC's." The ROC's are (p. 104):

1. Subjects must understand, be familiar with, the commodity to be valued.
2. Subjects must have had (or be allowed to obtain) prior valuation and public choice experience with respect to consumption levels of the commodity.
3. There must be little uncertainty.
4. WTP [willingness to pay], not WTA [compensation demanded], must be elicited.

Cummings and his co-authors argue (p. 104) that, "The relevance of the ROC's lies in our expectation that, if the CVM [contingent valuation method] institution satisfies them, we would expect the resulting measure of value to approximate market-analogous values within a range of error defined by 'background' sources of error, suggested at the present time to be no less than ± 50 percent."

If the ROC's represent necessary and sufficient conditions, then they would certainly raise grave concerns about the use of contingent valuation to estimate non-use values. Those who

hold existence values and those who are potential future users with option values may lack the direct contact with the affected resource required to fulfill ROC 1. Subjects asked to express non-use values are likely to lack the prior experience necessary to satisfy ROC 2. Option values, by definition, involve uncertainty. In addition, existence values may be held in the presence of substantial uncertainties (e.g., the full extent and irreversibility of effects and the potential for successful restoration). Hence, non-use value studies may violate ROC 3. And, as argued above, some would argue that compensation demanded is the appropriate measure of natural resource damages, yet ROC 4 would rule out measuring compensation demanded with contingent valuation techniques.

The ROC's are useful here in focusing our attention on kinds of misgivings that many economists have about applying contingent valuation to non-use values. More recently, cognitive psychologists have begun to express misgivings as well. Many of their concerns have focused on what has come to be known as "embedding effects" and other such phenomena.

Embedding and Related Phenomena

Kahneman and Knetsch (forthcoming) deal with embedding and what they term an "ordering effect" at the same time. They present the problem in this way (pp. 3-4):

> The standard interpretation of CVM results is that the aggregate WTP for a good is a measure of the economic value associated with that good, which is fully comparable to values derived from market exchanges. However, two related observations that cast doubt on this interpretation have been discussed in the CVM literature. The first is an order effect in WTP responses when the values of several goods are elicited in succession: the same good elicits a higher WTP if it is first in the list than if it is valued after others. . . . Because the order in which goods are mentioned in a survey is purely arbitrary, any effect of this variable raises questions about the validity of responses.
>
> Another problem for CVM is an effect we shall call <u>embedding</u> . . . : the same good is assigned a lower value if WTP for it is inferred from WTP for a more inclusive good than if the particular good is evaluated on its own. [emphasis in original]

According to Kahneman and Knetsch, ordering and embedding effects raise very serious concerns about the validity of contingent valuation. They argue (p. 6):

> The problem for interpretation of CVM results is the following: If the value of a given landmark is much larger when it is evaluated on its own than when it is evaluated as part of a more inclusive package of public goods, which measure is the correct one? The discussions of the problem in the literature provide no agreed principles that would define the proper level of aggregation for the evaluation of a specific good. In the absence of such principles, the results of CVM become arbitrary. This criticism could be fatal. No measuring instrument can be taken seriously if its permitted range of applications yields drastically different measures of the same object.

Though Kahneman and Knetsch are a bit hard to follow, they appear to have two concerns about contingent valuation. First, ordering and embedding effects raise, in their minds, what we shall term theoretical concerns. As a matter of principle, a measuring tool that gives different values for the same thing is theoretically unacceptable. Second, they appear to have concerns that are more empirical in nature. If people respond to contingent valuation questions so as to imply that one specific item (e.g., a scenic landmark) is worth almost as much as the whole class of goods (all scenic landmarks), then surely, according to Kahneman and Knetsch, contingent valuation is somehow failing to give valid value measures. They give an empirical example from Kahneman's (1986) own research where one sample of Ontario's citizens expressed willingness to pay for cleanup to preserve fishing in one small area of the province that was almost as large as another sample gave for cleanup of the waters in the province as a whole. Surely, they would argue, this indicated that contingent valuation failed to produce valid value estimates.

Though less forceful in their arguments than Kahneman and Knetsch, other writers have raised similar points. For example, Cicchetti and Peck (1989, p. 9) describe the phenomenon by asking,

> . . . how many of us would give the same response to a series of questions emphasizing hazardous substances in the environment as we would give at the same time, we were also asked about the nation's drug problem, finding a cure for AIDS, cancer or heart disease, achieving world peace, etc.?

Thus, the value of avoiding spills may vary depending on whether the drug programs, research on dreaded diseases, etc., are valued at the same time. Cicchetti and Peck do not say whether or not they believe that multiple values are valid causes for concerns about the theoretical or empirical validity of contingent values, but they do emphasize that the possibility that multiple values would give defense attorneys in Superfund cases a definite advantage.

By way of summary, then, non-use values measured using the contingent valuation method are likely to be viewed with great skepticism by at least some economists and other social scientists, particularly cognitive psychologists. However, such strong views are far from universal. We next consider some criticisms that have been leveled at the ROC's, then turn to empirical evidence regarding the validity of contingent valuation that raise further concerns about whether the ROC's are really binding. Following that, we shall take a closer look at the evidence on and interpretation of embedding and related effects.

Concerns About The ROC's

Randall and Kriesel (1990) reject the ROC's as either necessary or sufficient conditions for accurate contingent valuation. As sufficient conditions, the ROC's are to be rejected because, quoting Randall and Kriesel (p. 173), "it would be a simple task to devise invalid CVM formats that nevertheless conform to the ROC's." As necessary conditions, their arguments against the ROC's are based on the fact that citizens, for example in their voting behavior, react to real world situations that have not been experienced previously and that may involve substantial uncertainty. Is it reasonable to assume that they will be totally incapable of doing so in a contingent valuation exercise?

Mitchell and Carson (1989, pp. 92-94) also reject the ROC's as potential guidelines for determining when contingent valuation is applicable and when it is not. They point out that the ROC's are based on an idealized market model of consumer choices. Frequently, such ideals are not realized in real markets, even when expensive purchases are made. "Market researchers have long recognized that many purchases are infrequently made and that the information people gather before making purchasing decisions differs greatly, depending on the purchase

situation, the type of good, and the consumer's past experience" (Mitchell and Carson, 1989, p. 93). If consumers can function in real markets that deviate from the ideal, then why assume that they cannot predict with reasonable accuracy how they would behave in hypothetical markets that vary from the ideal? Mitchell and Carson (1989, p. 93) go on to argue that political "markets," especially referenda, may be a more relevant and appropriate model for exploring preferences for public goods than the private goods market model that provides the basis for the ROC's. They stress (p. 93), "The strict application of a private goods market model ignores any but self-interested consumption behavior and therefore downplays the 'public-regardingness' behind existence values. These values can affect political behavior." They go on to cite results from studies that show the superiority of political institutions, compared to private markets, in reflecting broader, more public-spirited preferences of citizen. Much of their book is devoted to developing the referendum format for contingent valuation, a format that they argue works well even under conditions that would violate the ROC's.

Thus, while the ROC's represent the thinking of one group of economists, the cited material from Randall and Kriesel (1990) and Mitchell and Carson (1989) represents the views of a second group who believes that well-designed contingent valuation studies can be successfully applied to a wide range of problems including measurement of non-use components of total values. As we pointed out earlier and will discuss in detail in the next section, there is a growing body of empirical evidence that supports the second group.

Empirical Evidence on the Validity of Contingent Valuation

How accurate is contingent valuation? Broad generalizations are not justified in answering this question. Within the range of primary data collection methods used in the various sciences, surveys are relatively easy and inexpensive. However, those untrained in survey design and administration often underestimate the difficulty and expense of implementing a survey. As a result, the contingent valuation field has attracted more than its share of the incompetent and poorly trained, as well as a few outright hacks. Many respondents will try to answer any question addressed to them, even if it does not make much sense to them. Poorly designed studies and studies dealing with particularly difficult situations could yield very

inaccurate results. As a consequence, there are some very poor studies in the literature, especially the grey literature. We would be the last to try to defend contingent valuation studies across the board. Still there is evidence, based on laboratory and field experiments where contingent values can be compared to values based on actual transactions, indicating that contingent valuation can produce relatively accurate values for willingness to pay.

Consider, first, our own work involving hunting permits. In three different experiments, we bought and sold permits to hunt either Canada geese or white-tailed deer. These studies are described in detail elsewhere (Bishop and Heberlein, 1979; Bishop et al., 1983; Welsh, 1986; Heberlein and Bishop, 1986; Bishop et al., 1988; Bishop and Heberlein, 1990). By way of summary, contingent measures of willingness to pay performed rather well. For example, one measure of contingent willingness to pay for a permit for a special deer hunt in the Sandhill Wildlife Area was $35 while a comparable value based on actual cash transactions was $31, and the values were not statistically different. Nearly all the hunters in this study had little or no previous experience hunting in Sandhill, although many were experienced deer hunters. Thus, one might question the case for ROC 1, which requires that respondents in contingent valuation studies be familiar with the item being valued. At least, one would want to ask what degree of experience and familiarity is really required. ROC 1 (and ROC 2 for that matter) is very vague in this regard. Furthermore, the permit involved here was for a one-day hunt. Hunting, it should be noted, particularly a one-day deer hunt, is subject to great uncertainty. In relatively unfamiliar terrain, which strategy would be likely to lead to success? Will it rain or snow on the one day of the hunt? What sorts of help or interference will come from other hunters? How will the deer move? Will a good shooting opportunity present itself? If so, will the deer involved be ordinary or of trophy quality? Despite this uncertainty, contingent valuation measures of willingness to pay did well, raising doubts about ROC 3, which would limit contingent valuation applications to cases where there is little or no uncertainty. On the other hand, contingent willingness-to-accept compensation displayed a strong, statistically significant tendency to overvalue the permits, a result supporting ROC 4.

A second field experiment involved the sale of fresh strawberries (Dickie, Fisher, and Gerking, 1987). This experiment involved personal interviews carried out at the houses of residents of Laramie, Wyoming. One sample of residents was given an opportunity to actually purchase strawberries. A second sample was asked for hypothetical bids for strawberries. The

researchers found no statistical difference between a demand function estimated from data based on actual transactions and a demand function based on hypothetical data. This is a study where the ROC's came close to being fully fulfilled and contingent valuation did perform very well. It raises doubt about the conclusion of Cummings et al. that, at best, one can hope for accuracy no better than ± 50 percent. In the strawberry experiment--and the hunting permit experiment for that matter--contingent valuation did much better than that.

A laboratory experiment was conducted at the University of Wyoming using undergraduates as subjects (Coursey, Hovis, and Schulze, 1987). An unpleasant environmental stimulus was simulated using a bitter, unpleasant, but harmless substance, sucrose octa acetate (SOA). The students were randomly assigned to groups that would either focus on willingness to pay or compensation demanded. Willingness-to-pay groups and compensation-demanded groups were kept entirely separate throughout the study. The experiment was conducted in three parts. In the first part of the experiment, students were provided with verbal descriptions of SOA and asked in a contingent valuation format to state either how much they would be willing to pay (hypothetically) to avoid drinking a small amount or how much they would have to be paid (hypothetically) to do so. The second part of the experiment involved three steps. First, each subject tasted a few drops of SOA. Then each was asked for his or her revised willingness to pay or compensation demanded. In the final step, the researchers attempted to bid up the members of the willingness-to-pay group and bid down the compensation-demanded group in 25-cent increments. Thus the first two parts of this experiment provided estimates of willingness to pay and willingness to accept compensation collected by using several contingent valuation techniques. In the third part of the experiment, subgroups of eight individuals each participated in an auction designed to elicit actual cash bids. At the end of the bidding, those in willingness-to-pay groups who were the high bidders paid and those who were not drank the SOA. Those in compensation-demanded groups who won the auction with minimal bids were actually paid and drank some SOA.

On the willingness-to-pay side, bids in all three parts of the experiment were quite close, indicating that contingent valuation performed well. Given that SOA was completely unfamiliar to the participants at the beginning of the experiment, this raises additional questions about the empirical basis for ROC's 1 and 2. That respondents were in general familiar with tasting substances and finding some of them unpleasant was apparently sufficient to allow them to

express accurate contingent willingness-to-pay values. This is all the more interesting given that one of the authors of this study, William Schulze, was also a co-author of the ROC's. Initial bids for compensation demanded were, on average, quite high relative to willingness to pay, but by the end of the third part of the experiment, mean compensation demanded had collapsed and become statistically indistinguishable from mean willingness to pay. Thus this experiment supports ROC 4 and casts further doubt on the ability of contingent valuation mechanisms to measure compensation demanded.

A similar experiment employed the same basic design as the SOA study but involved a pleasant tasting raspberry drink (Brookshire et al., 1988). Compared to the SOA experiment, contingent willingness to pay for the raspberry juice seemed to underestimate the actual cash value by a substantial amount. The mean bid in the first part of the experiment, which most closely approximated what would have occurred in a contingent valuation study, was $1.41, while the mean bid in the final round of the auction was $10.66. A preliminary judgment would be that contingent valuation did not work very well in this case. However, the results may have been unduly affected by aberrant behavior of several participants. For example, on the final round of the auction, one participant bid $100 and two others bid $50. Several other bids in the $5 to $10 range were reported. Such bids seem excessive for an eight ounce glass of juice, particularly from undergraduate students. Perusing the data provided in the Brookshire et al. (1988) paper, one wonders if social dynamics within the auction groups may have affected the results. Results for compensation demanded showed again that contingent valuation overestimated the value, supporting ROC 4.

Kealy et al. (1988) also performed a laboratory experiment involving undergraduates. In this case the commodity was a well-known brand of candy bar and only willingness-to-pay measures were collected. In the first step, subjects were randomly assigned to one of two groups, one group to engage in actual transactions and the other group to engage in hypothetical transactions. The "actual-transaction group" was told from the beginning that its members would have an opportunity at some point in the future to actually buy a candy bar. The "hypothetical-transaction group" was asked to imagine that such an opportunity would present itself in the future. During the first step, both groups participated in contingent valuation exercises including dichotomous choice questions, which asked whether they would buy one of the candy bars at a specified price, take it or leave it. The prices were varied across respondents. They were also

asked open-ended valuation questions requesting them to state their maximum willingness-to-pay amounts. Both groups were invited back two weeks later and were again asked open-ended contingent valuation questions for the candy bar. In addition, each member of the actual-transaction group was given the opportunity to buy a candy bar at the dichotomous-choice amount that each had responded to two weeks previously.

The contingent values for the candy bars were quite close across valuation methods (dichotomous choice and open-ended), across the actual-transaction and hypothetical-transaction groups and across time. Expected values for the candy bar were estimated as simple means of responses to the open-ended questions, while dichotomous-choice responses were analyzed using probit analysis and expected values were calculated from the resulting estimated cumulative probability density functions. The expected values ranged between $0.76 and $0.85, and the various contingent values were not statistically different. However, when the actual transactions group was offered the opportunity to complete a transaction at prices they had been asked about in the dichotomous-choice contingent valuation exercise, some respondents who had previously agreed to pay their respective amounts backed out. As a result, the expected value of the candy bar, based on actual transactions, was only $0.57, which was lower by a statistically significant amount compared to the contingent values.

In addition to results of such experiments, other types of evidence can be cited which casts some light on the validity of contingent valuation. Many concerns about contingent valuation have been based on fears that respondents will intentionally respond in ways designed to influence study results in desired directions. Such strategic behavior could involve either free riding (stating zero or very low values) to avoid perceived effects of truthful revelations on what they actually pay or overpledging, based on the perception that they can influence the supply of a desired public good and not have to pay for it. Both Hoehn and Randall (1985) and Mitchell and Carson (1989) have shown that even if respondents wish to behave strategically, they face complex, at times countervailing, incentives when attempting to design self-serving strategies. In other words, respondents who wish to respond strategically may have difficulty figuring out how to do so in their own best interests. Furthermore, a large number of laboratory studies have involved public goods and the possibility of strategic behavior. Relevant literature is cited in Mitchell and Carson (1989, Chapter 6) and Bishop and Heberlein (1990, p. 93). In general, strategic behavior has been difficult to document even in cases where it is a rather obvious

choice and seems to be much less pervasive than economic theory would lead one to expect. Though possible strategic responses must be considered in designing contingent valuation studies, the severe biases that early critics of contingent valuation feared are not a major threat to the approach.

Other researchers have compared value estimates derived using the contingent valuation method to values for the same item derived using other non-market valuation techniques. In general, the results of such comparisons have also supported the validity of contingent values. For example, contingent values have been shown to be roughly comparable to values derived from travel-cost models (Knetsch and Davis, 1966; Desvousges, Smith, and McGivney, 1983; Sellar, Stoll, and Chavas, 1985), hedonic price models (Brookshire et al., 1982; d'Arge and Shogren, 1989), and the prices of substitutes (Thayer, 1981).

In summary, the research results surveyed so far tend to support the accuracy of contingent valuation in measuring willingness to pay. The prediction of Cummings et al. (1986) that even under the best of circumstances the error in contingent values will be no less than ± 50 percent appears to be overly pessimistic. Contingent valuation performed better than that in the hunting permit, strawberry, SOA, and candy bar studies. The ROC's do not appear to be good guides to judging when contingent valuation will work well and when it will fail. The reason they are poor guides is that they are too vaguely defined. For example, what is really meant when they require that respondents be "familiar" with the good (ROC 1) and have had past valuation experience (ROC 2)? How much uncertainty is too much under ROC 3? Furthermore, the ROC's do not seem to be either necessary or sufficient to assure success. They appear to have been met in the strawberry and candy bar studies, yet contingent valuation appeared to perform well in the former and less well in the latter. The SOA and the hunting permit studies appear to have involved significant deviations from the ROC's, yet contingent values for willingness to pay proved to be rather accurate estimates of what respondents would really pay in both of these cases. Only ROC 4, requiring that willingness to pay--and not compensation demanded--should be elicited, is consistently supported by the empirical evidence.

Admittedly, so far we have addressed only studies involving use values. Many have questioned whether conclusions based on use values can be generalized to non-use values.

Evidence regarding the accuracy of contingent non-use values is still fragmentary, but it does provide some tentative insights.

Kealy et al. (1987) conducted a study of the value of reduced acid rain damage in the Adirondack Mountains that closely paralleled their study involving candy bars cited above. Again, student subjects were involved in a two-step experiment. In the first step, an "actual-transaction" group was told that each of its members would have an opportunity in the future to make a donation to reduce damages to aquatic ecosystems from acid rain and each was asked whether she or he would be willing to donate a predetermined, randomly assigned amount. At the same time, members of a "hypothetical-transactions" group were asked to imagine that such an opportunity to donate would be presented to them and asked whether they would donate specified amounts. Two weeks later, both were asked an additional contingent valuation question and each member of the actual-transaction group was given the opportunity to make an actual donation of the amount that had been specified two weeks before. Kealy et al. point out that some of their subjects were past and/or potential future users, so that some use values were probably involved, but they concluded that existence values as well as option prices were being expressed by many of their respondents.

As in the candy bar experiment, contingent valuation tended to overestimate the actual willingness of the students to donate. The actual-transaction group expressed an average value of $6.83 in the first step and revealed an actual willingness to donate $5.37, on average, in the second, a difference that was found to be statistically significant. The hypothetical-transaction group expressed a willingness to donate $15.51, on average, in the first step, and $10.11 in the second step, a difference that was statistically significant. Furthermore, the differences between the actual-transaction and hypothetical-transaction groups' willingness to donate are statistically significant at both steps. Contingent willingness to donate exceeded actual willingness to donate by a factor of 2.3.

A second experiment of direct interest here is Boyce et al. (1989), discussed above. After carrying out the contingent valuation exercise described above, the researchers allowed the four groups of participants to participate in a real auction. Actual transactions were carried out based on the results of the auction.

Recall that in the contingent valuation exercise dealing with compensation demanded, respondents showed a strong reluctance to participate, especially when the purchased trees were to be killed. For that group, 14 of the total of 29 participants refused to participate in the contingent valuation exercise and one additional person stated a bid of $5,000. Even when nothing was said about the fate of the trees, four people would not state a compensation-demanded bid in the contingent valuation exercise. Means reported here for contingent compensation demanded were calculated excluding the refusals (rather than, say, counting them at zero or infinity) and the $5,000 bid. Interestingly, such behavior did not carry over to the actual auction. There everyone participated, 29 in the group that was told that sold trees would be killed and 30 in the group that was told nothing about the fate of the trees.

Table 2 shows the mean values from the four parts of the experiment. Several conclusions seem to be implied. First, whether the surplus trees would be killed made a substantial difference in the auction bids. On the willingness-to-pay side, such knowledge increased the mean bid by a factor of 1.6. For willingness-to-accept, the value was 2.3 times larger if the trees were going to be killed. The authors concluded that this is evidence of the presence of existence values. They go on to argue that such ratios are likely to be highly dependent on the specific circumstances surrounding value formation. They say (p. 333):

> We hypothesize that the following characteristics would tend to increase existence value: (1) uniqueness of the good, (2) time to become familiar with the good, (3) a living rather than nonliving good, (4) level of personal involvement with the good, (5) the violence with which the good would cease to exist, (6) the willfulness with which the good would cease to exist, and (7) the degree of control the valuator has over the fate of the good.

Table 2: Results from the Norfolk Island Pine Experiment

	CV	Auction
Willingness to Pay		
Kill trees	$16.80	$7.81
Fate of tree not mentioned	$ 6.06	$4.81
Compensation Demanded		
Kill trees	$26.07	$18.43
Fate of trees not mentioned	$14.12	$ 8.00

Source: Boyce et al., 1989

They suggested that if these hypotheses are valid, then their Norfolk Island Pines had lower existence values than they would have had if they been unique and more familiar to respondents (for example, if respondents were "introduced" to their trees at the beginning of the experiment). On the other hand, that the trees were living things and that they were placed in front of the participants when they arrived--rather than simply being described to them--tended to increase the existence value. Perhaps the fact that the researchers clearly specified that the plants would be destroyed also increased the value compar to a less violent end for the trees. Willful destruction, under hypothesis (6), may tend to stimulate higher values compared to accidental loss or natural processes such as death due to cold weather. Finally, subjects had complete control over the fate of the plants, which meant that they bore complete responsibility for what happened to them. Bids might have been lower if respondents had not been placed in such an extreme position.

Another interesting result from the auction was a large disparity between willingness to pay and compensation demanded that persisted despite the fact that subjects participated in repeated trial auctions before final bids were taken. This is a marked contrast to the SOA and raspberry juice experiments discussed above, where repeated trials resulted in substantial decreases in willingness to accept by the time final bids were taken. In the tree experiment, the

auction mechanism involved recording bids on a computer before a ball indicating the price was drawn. The SOA and raspberry juice experiments involved groups of eight subjects participating in a Vickrey auction. On the willingness-to-pay side, for example, sealed bids were taken and the four highest bids were accepted with the winners all paying the highest unsuccessful bid. While theoretically incentive compatible, the Vickrey auction with such a small group may have stimulated dynamic social processes within the groups that affected the outcomes. We have already noted such a possibility for the raspberry juice experiment and similar processes may have affected the SOA experiment as well.

Finally, contingent valuation tended to overvalue the trees in all cases. On the willingness-to-pay side, this tendency may have been stronger when the trees were to be killed than when they were not. Unfortunately, the Boyce et al. report to which we had access did not provide either statistical test of significance or the statistics required for the reader to perform such tests. We strongly suspect that the difference between the contingent and auction values for willingness to pay is statistically significant for the treatment where the unsold trees would be killed, while it may not be for the treatment where the fate of the trees was not discussed. The high rate of non-participation in the contingent compensation-demanded treatment (in which the trees purchased by the researchers were to be killed) clouds any attempt to make comparisons between contingent and auction values there. We would simply conclude that Boyce et al. provides further support for ROC 4.

One other study, and one of a very different sort, is relevant in the present context. Carson, Hanemann, and Mitchell (1986) and Mitchell and Carson (1989, pp. 203 and 204) describe a study constructed around an actual California referendum, Proposition 25 on the November 1984 ballot. If passed, Proposition 25 (entitled the Clean Water Bond Law of 1984) would authorize issuance of 20-year bonds totaling $325 million, mostly for construction of sewage treatment plants. A contingent valuation exercise was carried out in which each respondent was asked how he or she would vote on the issue if passage of the issue resulted in alternative annual costs to them. The annual costs presented to each respondent ranged from $1 per year to $50 per year. Before the election, but after the survey, each registered voter received the California voter's pamphlet, which included information on the monetary commitment that adopting Proposition 25 would entail. Based on responses to the contingent valuation question and the annual cost reported in the voter's pamphlet, Carson, Hanemann,

and Mitchell predicted that the measure would pass with between 70 and 75 percent of the vote, depending on the voting behavior of those who were undecided about their voting intentions at the time of contingent valuation survey. The measure passed with a vote of 73 percent in favor. This result, as well as other studies of voting cited by Mitchell and Carson, seems to support the validity of contingent values using referendum formats. Based on the rather consistent evidence from contingent valuation studies to be reviewed later in this report, we would infer that respondents in the Carson et al. study were motivated by non-use and use values.

Implications

Before addressing the issue of embedding it will be useful to draw some tentative conclusions about the accuracy of the contingent valuation method as revealed by the studies just reviewed.

First, contingent measures of compensation demanded do not appear to have much validity either for use or non-use values. Contingent compensation demanded appears to be strongly biased toward overestimates of value. It also may tend to elicit refusals to participate, particularly where emotional issues are involved, as in the tree experiment. This is likely to be a troublesome problem in studies assessing non-use values of natural resource damages, especially if living resources are involved. Experimental results indicate that compensation-demanded values may also be subject to greater instability over repeated trials than willingness-to-pay values. This is important because contingent valuation studies typically involve only one opportunity for a response, rather than the feedback and learning that can come from repeated trials.

Second, contingent willingness to pay seems to provide reasonably accurate measures of use values. While economists continue to haggle about the fine points, a well-designed contingent valuation study is probably no more subject to error than a well-designed travel-cost or hedonic-price study. Contingent values are probably superior to unit day values and other such values based on imperfect market substitutes. There is sufficient evidence to expect that

contingent use values for willingness to pay can be treated as lower bounds for actual compensation demanded.

Third, while a healthy degree of skepticism is warranted, the successes of contingent valuation in measuring use values should be viewed as prima facie grounds for cautious optimism about the potential for measuring non-use values. The success of contingent valuation ultimately depends on the ability and willingness of respondents to provide the information requested. The basic assumption of contingent valuation research is that if respondents are given contingent valuation scenarios that they can understand and relate to their personal economic circumstances, they can make reasonably accurate predictions about how they would value the item in question. Though a great deal more research is needed, there is a growing body of evidence to support this assumption, as summarized here. If respondents can do this for use values, then why not for non-use values as well? More research along the lines of the experiments involving existence values summarized above is badly needed to investigate the validity of contingent non-use values. In the meantime, results from the acid rain and tree experiments, and probably from the study based on the Clean Water referendum in California, seem to show that existence values are real and that they can be measured, at least roughly, using contingent valuation.

Before turning to a discussion of what a study of non-use values in the context of the Glen Canyon Environmental Studies would look like, let us return for a moment to the concerns about contingent valuation based on embedding. Are there grounds in the Kahneman and Knetsch paper for changing our conclusions that non-use values can be measured at least roughly using the contingent valuation method?

What About Embedding?

Kahneman and Knetsch used data collected during a study of disaster preparedness in one region of British Columbia to illustrate the effects of embedding and the problems that embedding effects raise for contingent valuation. They found three different values for

increased availability of rescue-related equipment and trained personnel depending on the degree of embedding.

The lowest values they obtained (a mean of $13.57 and a median of $1) were for a subsample asked to state a value in the context of a high degree of embedding. The first step leading up to these values was to ask this subsample about their values for a broad range of environmental services, including (p. 7) "preserving wilderness areas, protecting wildlife, providing parks, preparing for disasters, controlling air pollution, insuring water quality, and routine treatment and disposal of industrial wastes." Then respondents were asked (p. 7):

> If you could make sure that extra money collected would lead to significant improvements, what is the most you would be willing to pay each year through higher taxes, prices, or user fees, to go into a special fund to improve environmental services?

Next, all respondents in this subsample who gave non-zero responses were asked (p. 7):

> Keeping in mind the services just mentioned, including those related to providing parks, pollution control, preservation of wilderness and wildlife, and disposal of industrial wastes, I would like to ask you in particular about improved preparedness for disasters. What part of the total amount that you just mentioned for all environmental services do you think should go specifically to improve preparedness for disasters?

Then, those who gave non-zero responses in the second level were reminded about the various aspects of disaster preparedness such as emergency services in hospitals, maintenance of large stocks of medical supplies, food, etc. Then they were asked (p. 7):

> Keeping in mind all aspects of preparedness for disasters, what part of the total amount you allocated to improving preparedness do you think should go specifically to improve the availability of equipment and trained personnel for rescue operations?

Thus, for this subsample, the value of improvements in equipment and trained personnel was deeply embedded in the value of an array of environmental services and the full range of services needed for disaster preparedness. The result was the mean value of $13.57 mentioned above.

A second subsample participated in a similar exercise, except that the initial question they answered referred to "a special fund to improve preparedness for disasters," with a subsequent allocation to "go specifically to improve the availability of equipment and trained personnel for rescue operations" (p. 8). Here the mean value of equipment and personnel was $76.65, with the median of $16.00.

Finally, a third subsample was asked directly, without embedding, about the value they would place on increased availability of rescue equipment and personnel. Here the mean was $122.64 and the median was $25.00. Thus the mean increased by a factor of nine and the median by a factor of 25 between the unembedded and the deeply embedded designs.

Equally disconcerting was the closeness of the values at the first stage of the different question designs. The environmental improvements were worth $135.91 to the first subsample at the beginning of its contingent valuation exercise. Improved disaster services were worth $151.60 to the second subsample when they began at that level. And, for the third group, which valued only increased personnel and equipment, their value, as already noted, was $122.64. The differences between these figures were not statistically significant.

Kahneman and Knetsch argue that these results are implausible as representations of true values and cannot be explained away by economic theories of substitution or income effects. We agree but would argue that these results do not represent a valid empirical test of embedding effects because of flaws in study design.

There is broad agreement among contingent valuation researchers that, for results to be valid, respondents must understand in some detail what they are being asked to evaluate and under what circumstances. In fact, this principle has recently been articulated well by a pair of non-economists, Fischhoff and Furby (1988). Fischhoff and Furby offer economists some sound advice about the psychology of transactions as it relates to contingent valuation. They point out (p. 151, emphasis in original):

> Any proposed transaction has three constituents: something being received, something being given in exchange, and a social context within which the exchange would be enacted. In an economic transaction, these might be called the good, the payment, and the marketplace. . . . For a transaction to be satisfactory, each of these three constituents must be well defined and well understood by all the participants.

They explore in some detail the characteristics of each of these constituents of a satisfactory transaction. With respect to the good, they say (p. 153, emphasis in original):

> Although they are transferred as wholes, goods may be thought of as bundles of attributes, representing outcomes of accepting the transaction that might be valued either positively or negatively. The first step in defining a good is identifying its potentially valued attributes.

They point out that (p. 159, emphasis in original):

> Any transaction involves a change of state. . . . We use reference level for the state obtained if the transaction is not enacted and target level for the state obtained if it is. . . . Reference and target levels must be specified for every (potentially) valued attribute affected by the transaction.

Fischhoff and Furby go on to discuss relevant aspects of the comparison between the reference and target levels of the good's attributes, mentioning the need to be clear on the extent of the changes, their timing, and the certainty of provision. They develop, in a similar way, the need to convey to people in clear and complete terms the nature of payments and the social context of proposed transactions. They summarize by saying that (p. 179), "Specifying all relevant features [of transactions], and ensuring that they have been understood, is essential to staging transactions."

Though they bring a fresh perspective, most of what Fischhoff and Furby say is consistent with the economics literature on contingent valuation survey design. For example, Randall et al. (1983, p. 637, emphasis added) have stated:

> Contingent valuation devices involve asking individuals, in survey or experimental settings, to reveal their personal valuations of increments (or decrements) in unpriced goods by using contingent markets. These markets define the good or amenity of interest, the status quo level of provision and the offered increment or decrement therein, the institutional structure under which the good is to be provided, the method of payment, and (implicitly or explicitly) the decision rule which determines whether to implement the offered program. Contingent markets are highly structured, to confront respondents with well-defined situations and to elicit a circumstantial choice contingent upon the occurrence of the posited situation.

To take just one more instance, Bishop and Heberlein (1990, p. 83) pointed out that:

> If contingent values are to be valid and useful, the object being valued must be appropriately defined to reflect the policy issue being addressed. Thus, good studies involve carefully presented descriptions of the resources or changes in environmental quality that are to be valued.

The need to clearly and completely describe the item to be valued, how payment will be made, and other potentially relevant details is spelled out repeatedly in such standard works as Mitchell and Carson (1989) and taken for granted by authors such as Cummings et al. (1986). The principle is simple and almost too obvious to need stating: People cannot value something if they do not understand what it is or the terms of the deal.

In all three treatments for evaluating improvements in the availability of disaster-related equipment and personnel, respondents were asked to evaluate vaguely defined products under vaguely defined terms. Whether the valuation exercise is focusing on the environment, disaster preparedness in general, or personnel and equipment in particular, the reference level of provision is not described at all and the target levels are merely described as "improvements." Nothing is said about which particular attributes will be improved, about the physical location of the changes, about the timing of improvements, or other potentially relevant aspects. Instead of specific details about the proposed transaction, respondent are left with vague references to taxes, prices, and user fees to be placed in some undefined "special fund."

Kahneman and Knetsch do present, in summary form, results from other studies they have concluded that seem to further justify their concerns. Few study design details are

included, so it is difficult to evaluate the other studies in the same way as we have done for the study of disaster preparedness. From the information given, it appears likely that the other studies suffered f . similar flaws. Our conclusion is that the Kahneman and Knetsch paper is not based on state-of-the-art contingent valuation procedures. In fact, if one wanted to design a contingent valuation study that would fail, the disaster preparedness study would be a good choice to use as a model. Accordingly, though we anticipate that while their paper will touch off a long and arduous debate, the Kahneman and Knetsch results do not represent an adequate base to challenge the validity of contingent valuation.

Based on the experimental and value comparison studies that we have attempted to summarize in this report, we believe the assessment of Mitchell and Carson (1989) has a much stronger empirical justification. They addressed somewhat the same set of empirical concerns as Kahneman and Knetsch under the heading of "amenity misspecification bias" (p. 249):

> Since people tend not to have previously well-defined values for many of the goods valued in CV studies, there is considerable potential for them to ignore some or all of the details in a scenario, or to distort them by unconscious use of judgmental heuristics.

Two examples of amenity misspecification bias are symbolic bias and part-whole bias. Continuing to quote them directly (p. 249-251, emphasis in original):

> Symbolic bias occurs when respondents react to an amenity's general symbolic meaning instead of to the specific levels of provision described [in the contingent valuation question]. . .
>
> Part-whole biases are major amenity misspecifications, and are also a result of the tendency of respondents to respond to public goods as global symbols without paying sufficient attention to the specific description offered in a CV survey. The dimensions of a good that are particularly prone to this misperception are its geographic distribution, its benefit composition, and the package of policies of which it is a part. Consider a respondent who is asked how much she is willing to pay for water quality improvements in a local river basin. If she is unable to isolate that river basin in her mind from her state's or region's other rivers, the respondent may in fact value a larger range of waters than intended by the researcher.

In this way, Mitchell and Carson would seem to propose an explanation for the empirical part of Kahneman and Knetsch's concerns. If people give very similar values for one scenic landmark and all such landmarks, it could be because they have not really come to grips with their values for the single landmark but are expressing more general concerns for all scenic landmarks in both cases. However, Mitchell and Carson are much less pessimistic than Kahneman and Knetsch about the ultimate implications of such problems for the validity of contingent valuation studies, suggesting that there are promising strategies for survey design that can minimize the tendency of respondents to confuse the part with the whole. Thus, although the Kahneman-Knetsch results warn of major pitfalls that poorly designed studies can encounter, pitfalls that must be considered in designing satisfactory studies, they hardly justify rejecting the method.

SECTION 3: A PRELIMINARY RESEARCH DESIGN FOR TOTAL VALUATION OF COLORADO RIVER RESOURCES

This section presents a preliminary research plan for a total valuation study under GCES. As the first two sections of this report have amply demonstrated, this is a very new area for resource economists. There are no standard responses to many of the questions that must be answered in designing such a study. Hence, it is impossible to lay out in great detail all of the steps that would be taken to reach final results. Instead, the study plan described in this section includes two decision points at which progress will need to be reviewed and decisions made as to whether the research should be modified or terminated in light of what has been learned.

A three-stage project is envisioned, with decision nodes after the first and second stages are completed. The first stage will consist of a program of qualitative research designed to address two issues: (a) Is it feasible to describe the consequences of changes in dam operations in terms that typical survey respondents can understand? (b) Do non-users have preferences with respect to the kinds of environmental changes that occur as a result of dam operations? The goal of the qualitative research is to gain some understanding in broad terms about how people think about the Colorado River resources studied under GCES. As opposed to the first stage, the second and third stages will consist of quantitative research. If the results of the qualitative research demonstrate that non-users as well as users believe that they have a stake in the impacts of dam operations and that it is feasible to describe these impacts in a survey, then the second stage in the research will be to conduct an attitude survey. The purpose of this attitude survey would be to test hypotheses generated during the qualitative research program. We anticipate that this attitude survey would focus primarily on quantifying (in a non-economic way) the manner and degree to which various populations care about the resources that are affected by operations of Glen Canyon Dam. The third stage of the research, if carried out, would involve a contingent valuation survey. The purpose of the contingent valuation survey would be to quantify in economic terms using a total value framework, the impacts of various operations of Glen Canyon Dam.

General Research Issues

Several general research issues will be common to some or all three stages of research. In this next section we will describe what we consider to be the key issues that will have to be addressed as a research design is finalized and implemented.

Alternatives Versus Response Surfaces. It is worth emphasizing that the ultimate goal of the research described in this section is not to value dam operations per se. Rather, the objective is to value changes in the energy and environmental resource service flows that are caused by changes in dam operations. This is a crucial issue because it may ultimately affect the design of both the attitude and contingent valuation surveys that are planned for the third stage of this research. For example, suppose that it is ultimately determined through the qualitative research and the attitude survey described below that some resources affected by dam operations are relevant to non-use values. There are two ways in which the impacts to these resources could be presented to survey respondents in a contingent valuation survey.

One approach is to describe the impacts to all relevant affected resources on an alternative by alternative basis. One reason for using this approach is that any particular dam operation will affect all of the relevant resources in some way. If the impacts to resource flows occur in a bundled way, then it makes sense to describe the impacts in the same way. Using this approach also has the advantage of focusing the research on attaching values to a limited set of resource outcomes, namely, the resource outcomes predicted for the dam operation alternatives being evaluated. There is, however, a danger in following this first approach. Suppose that after completion of a contingent valuation survey, additional biological and physical research revealed that the descriptions of the outcomes used in the valuation survey were substantially different from the actual outcomes. In such a case, it would be very difficult, if not impossible, to say anything about the value of the actual outcomes. If the value of this new set of impacts was critical to the overall decision-making process, then it might be necessary to conduct a separate contingent valuation survey asking about this specific set of impacts.

A second approach to describing the impacts of dam operations avoids this problem. Under this second approach, respondents would be asked to evaluate impacts which vary in a

manner systematically determined by the researchers rather than the actual impacts of a specific set of dam operations. The ultimate objective of this second approach would be to estimate a function that maps the status of the relevant resources into an economic value. If n resources were determined to be relevant to non-use values, the estimated function would define an n+i dimensional valuation surface. Each point on the valuation surface would correspond to the value of a particular set of the levels of the relevant resources. To evaluate any particular dam operation alternative, all that would be required is a description of impacts of the alternative on the relevant resources. A major advantage of the valuation surface approach is that it would provide a great deal of flexibility to evaluate a much broader range of environmental changes than those directly associated with the proposed dam alternatives under consideration at the time the contingent valuation survey was implemented.

The valuation surface approach has at least one major drawback. The most useful valuation surface would be one that covers all of the possible resource outcomes. However, to achieve this goal, the research design would require that survey respondents evaluate a broad range of resource outcomes. Estimation of a "broad" valuation surface could require substantially more data collection expense than simply estimating the value of the resource impacts of a limited number of dam operation alternatives.

The review panel comments in the appendix strongly support the goal of deriving a valuation surface. In our opinion, while a valuation surface may represent the ideal research design from a purely theoretical point of view, whether estimation of such a valuation surface will be technically or practically feasible (and at a reasonable cost) remains to be determined. However, it is not crucial at this point to fully resolve this issue. Indeed, information collected to initiate the qualitative research and the results of the qualitative research itself should provide valuable insights into the relative merits of the two approaches.

Given the immediate need to provide information for the Glen Canyon Dam Environmental Impact Statement, we are initially proposing that the total valuation study be formulated around the alternatives that are to be formally considered in the EIS. This will give the study some necessary boundaries in terms of environmental and other effects to be evaluated. In terms of the ultimate goal of producing scientifically credible total values for downstream resources, a baseline is needed and we propose that the baseline be current

operations. Consequently, study participants at each stage will be asked about their attitudes and values for changes in environmental and energy services resulting from operational or structural changes from current dam operations.

Identifying and Describing Resource Impacts. Once the range of specific alternatives is defined, the downstream resources most likely to be affected by the alternatives must be identified. Then a summary of the resources that are thought to be affected by dam operations and the processes through which they are affected will be prepared. Verbal, written, and perhaps visual, descriptions based on this summary will form the basis of the information provided to respondents in both the equalistic and quantitative portions of the research. Where necessary, variations in impacts depending on hydrological conditions will be considered explicitly. It will be necessary to devote sufficient effort to summarize current knowledge in a form that GCES researchers will agree is fair, comprehensive and understandable. In preparing this summary we envision reviewing previous GCES research, and discussing the status of current research efforts with a handful of GCES researchers. Next, we will prepare a written description of the impacts of current dam operations and the alternatives and send it to select GCES researchers for review. This description will be revised from time to time as new results become available.

The potentially affected set of resources is, up to a point, obvious from previous work on GCES. Very likely candidates include beaches, endangered fishes, other native fishes, cultural sites, rare and endangered birds (peregrine, southern bald eagle), the richness of the Grand Canyon riparian ecosystem in terms of flora and fauna, and recreational opportunities for white-water boating and trout fishing. What we have said about the motives for existence values indicates that the generating resources of Glen Canyon Power Plant should be included in this list. That is to say, if total values are motivated at least in part by altruism, individuals may suffer economic losses if the cost of electricity rises for those who consume power produced at Glen Canyon Dam. Consequently, to the extent that alternatives might affect the availability or cost of power to consumers, the effects could have existence values. Other resources to be considered at this level will no doubt be identified.

During the early stages of the research process it is important to be as comprehensive as possible in defining the set of potentially affected resources. On the other hand, describing the

conditions of a large set of resources under both the baseline and the alternative conditions in terms the general public can understand could be a formidable (if not impossible) task. It may be necessary to reduce the number of resources considered in order to simplify the information that must be conveyed to participants in the contingent valuation survey. If so, this reduction would be based on the results of the qualitative research and the attitude survey.

Fortunately, it is not clear that the full list of resources would be relevant for total values. It seems likely that the humpback chub would be included; although, even here verification is needed. The beaches along the river corridor are likely to be an important component. Cultural sites at risk seem likely to be included. It is less clear whether the trout fishery, made up as it is of exotics to the ecosystem, would have substantial non-use values. It will be necessary to investigate whether non-users feel that they benefit from the trout fishery and why. Riparian birds found throughout the Southwest represent a similar case. Perhaps the fact that such birds are common means that they can be ignored for our purposes and perhaps not.

As mentioned above, power resources may also be included. However, many questions about how non-users view this resource remain to be answered. Does the fact that hydropower avoids air pollution and hazardous waste problems have some bearing here? Do altruistic motives underlying existence values extend to the electricity costs paid by others? The evidence collected during the focus groups--and ultimately in the attitude survey--is essential to answering these questions. Without this empirical evidence, such questions would lead only to speculation and debate, with the credibility of the final results suffering.

In this research, defining potentially important resource relationships is a critical step. Sound environmental economic analysis must be based on a sound scientific understanding of the underlying physical and biological effects of dam operations. To complete a total value study in time for inclusion in the EIS, we will need a fairly accurate understanding of what GCES, Phase II, will have concluded at the time the EIS is completed. Basing a total value study on scientific conclusions that are later modified substantially or failing to recognize aspects that later prove important would reduce the validity of the results. Scientists are often hesitant to state candidly what they expect to find prior to completing their analyses, and when they do make such statements, they may well change their minds later.

While this risk cannot be totally eliminated, some steps can be taken to minimize it. As noted above, the descriptions of the impacts of change in operational and structural alternatives on resource service flows will be written out so that the scientists whose views are being represented can review them in detail. This write-up will need to be carefully evaluated in order to appraise whether there are sufficient prospects for firm scientific conclusions to support the total valuation studies.

Assuming that scientific uncertainty will be at tolerable levels, written descriptions of the state of the science would be reviewed periodically with the help of scientists whose views are being represented to modify and update the descriptions as the EIS process, GCES, Phase II, and any other relevant research efforts proceed. Subject to the need to have results for the EIS, the contingent valuation survey could be postponed as long as is feasible in order to incorporate the most recent scientific information.

Relevant Populations. How widespread is non-user interest in riverine resources? The research must carefully address whether this subset of the resources of the Grand Canyon is of sufficient interest on a national basis to warrant a national sample for the total valuation study or whether a regional sample including the nine western states or even the southwestern states alone would be more appropriate. However, determining the extent of national interest in the Grand Canyon river corridor will not be a simple matter to unravel. The fact that many of the resources of interest lie within Grand Canyon National Park and other Park Service administered lands may enhance the value of riparian natural, cultural and recreational resources on a national basis. The fact that tribal lands are involved may also be relevant to the national public, given the importance of Native American concerns. At the same time, as we have stated previously, some would argue that national interest is focused on the large geological structure of the Grand Canyon and associated vistas, rather than the resources of the narrow river corridor at the bottom of the Grand Canyon. The degree to which these points of view are valid is an empirical question that must be addressed through the focus groups and attitude survey. However, this study will begin with the presumption that the Grand Canyon is a national resource and that the appropriate sampling frame for the total value study is the population of the United States.

Comprehension of Resource Impacts. A closely related set of issues revolve around the ability of respondents drawn from the national or even a regional population to understand the very specific impacts addressed by GCES and the EIS process. Some might argue that people with limited personal familiarity with the Grand Canyon may see things in terms that are too "black and white," believing that Grand Canyon resources beyond the river corridor are being affected by dam operations regardless of how well we tell them otherwise. This is another issue that needs to be addressed empirically. In our view, however, such concerns indicate unwarranted pessimism about the intellectual capabilities of the citizenry. If Americans really cannot think in terms of the subset of resources that are at issue here, then our ability to carry out a study of how total values might be affected by dam operations and possible structural alternatives would be severely limited if not completely eliminated.

Dealing with Uncertain Impacts. Finally, we return to the thorny problem of scientific uncertainty. To our knowledge, there are no studies available to help anticipate the extent to which increasing uncertainty about underlying physical and biological effects influences the reliability and validity of contingent values of environmental resources. It is well known that uncertainty does cause people in laboratory experiments to behave in ways that appear to be either self-contradictory or causes them to place unwarranted weight on low-probability events. Such results tend to raise doubts about the validity of contingent valuation results when uncertainty plays an important part in the scenarios, as we saw in the Reference Operating Conditions of Cummings et al. (1986). But such effects have not been verified, much less expressed in quantitative terms. The qualitative research in stage one will provide an opportunity to investigate the ability of respondents to come to grips with the scientific uncertainty associated with GCES results.

Qualitative Research Plan

We plan to rely on focus groups as the primary tool for the qualitative phase of the research because they have proven in past studies to be a cost-effective approach to addressing the kinds of issues just discussed. As Alan Randall points out in his comments in Appendix A, it is sometimes necessary to supplement focus group discussions with one-on-one interviews of

study participants and individual debriefings of focus group participants to gain an in-depth understanding about key issues. In this subsection, we will describe how the focus groups will be conducted, recognizing the possibility that some of these additional qualitative tools may be utilized.

Focus groups are a qualitative research tool commonly used to gain insights into the range of understanding, attitudes, perceptions, opinions and thought processes of participants with respect to a specific set of issues. Several issues must be addressed when designing a research program based on focus groups. These include the number of participants per focus group, the number of focus groups to be held, and the agenda for the focus groups.

Number of Participants. Focus group discussions are typically conducted with a relatively small number of participants. The appropriate number of participants is affected by the complexity of the topics to be discussed, the number of topics, and the depth of responses desired from each participant. Having too few participants increases the chance that a full range of opinions about any given topic will not be represented among the participants. In addition, it may be difficult to generate in-depth discussions if the group is too small. On the other hand, if the number of participants is too large, it is difficult to provide an opportunity for each of the participants to respond in depth to all of the topics of interest. It has been our experience that focus group discussion tends to most successful when there are between five and ten participants in each discussion.

Focus Group Agendas. The focus group agenda is used to provide the moderator with guidance on the issues to be raised during the session. The moderator can pursue additional topics that arise as the session proceeds. While formal agendas will be developed as part of the research process, its main elements are easily anticipated. Nearly all focus groups begin with time for introduction of the participants to each other, a brief explanation of the topics to be discussed, and a short discussion of the ground rules governing the discussion. Second, we envision asking participants to tell us what they think of when they think of the Grand Canyon. This may be physical (e.g., deep hole, place in Arizona), legal (e.g., National Park), or another label (e.g., wilderness, solitude, place they have always wanted to visit). We will probe to try to learn more about the role of the river and the associated riparian ecosystem in their thoughts about the Grand Canyon. This portion of the agenda will help us to understand how people

visualize and think about the attitude object (in this case, the resource services affected by dam operations). This understanding will be crucial to understanding how stimuli can be presented to survey respondents so that they can value the elements of interest for our study and only those elements.

Third, the moderator will try to draw people out about their knowledge of the Grand Canyon, Glen Canyon, and Glen Canyon Dam and whether any of them have ever been to Glen or Grand Canyons or the dam, or read about them, or seen TV programs about them. Part of the purpose of this portion of the agenda will be to reduce the chance that some members who know more will set themselves up as authorities at the very beginning of the session with the result that we do not get the opinions of the participants who might be less familiar with the resources in question. We will also try to avoid this problem by experimenting with statements to the effect that we are most interested in the ideas of those who have not been there and would never personally want to go there.

The fourth part of the focus group discussion will involve a brief description of the dam, followed by asking participants how they feel the dam might have affected downstream resources. This discussion will establish the starting point for providing information in the attitude and contingent valuation surveys.

Fifth, we will provide respondents with descriptive material relating to resource service flows under the baseline and dam operating and/or structural alternatives. At least in later focus groups, the degree of certainty associated with the description of these impacts will also be described. Participants would then be asked to respond to these impacts. The moderator will probe to see whether participants understand the impacts and the uncertainty surrounding them. They will be asked how they view the different impacts and why they feel that way. The amount and specificity of information could be varied among the focus groups. The moderator would probe to see what information is considered important and what is not. Careful attention will be given to discovering what kinds of information people feel they need in order to express their preferences. The moderator will also probe what people consider to be substitutes for and complements to the environmental resources in question.

The final agenda item will be to explore the prospects for translating participants' attitudes about changes in resource service flows into monetary values. Do they really care enough one way or the other to pay hard cash for what they visualize as improvements or to avoid outcomes they feel are worse than the baseline? Particularly for those who do not view themselves as potential future users of the environment, why are they willing to pay anything? What is their reaction to alternative payment vehicles? Can they relate well to a question that is formulated as a referendum or will some other format perform better? Do focus groups in different parts of the country respond to the valuation exercise differently?

The forgoing discussion represents what we feel would be a reasonable focus group agenda. However, as mentioned earlier, the agenda for any particular focus group may be changed if an important issue is raised during the discussion that was not included in the original agenda. Likewise, we anticipate that the agenda will evolve as we learn more about specific issues. We also expect that hypotheses will be formulated about the various issues associated with the study and tested in a preliminary way by including them in the agenda for later focus groups.

Location and Number of Focus Groups. Since focus groups involve a relatively small number of individuals participating in a dynamic group process, there is always a possibility that the interpretation of issues based on a single focus group will be affected by some unique characteristic of the participants. On the other hand, it has been our experience that with a given agenda, very little information is gained from the third or fourth focus group with members drawn from a particular population. Usually we recommend that two focus groups be conducted with each population of interest because while the second group usually reinforces the results from the first focus group, this is not always the case. The second group can normally be held at a relatively low cost. A substantial part of the expense of conducting a focus group is associated with the preparation of an agenda and travel to the focus group site. These expenses are the same for a given site regardless of how many focus groups are conducted at the site.

We are proposing to hold the first two focus groups in the conference room at HBRS offices in Madison, Wisconsin. We will recruit the participants for these groups from a random sample of households in the Madison area. After completion of the Madison focus groups, we

will prepare a short report summarizing the focus group discussions, assessing the desirability of conducting additional focus groups, and suggesting revisions to the agenda, if necessary. Assuming the results of the first two focus groups in Madison are favorable, we propose to initiate a series of focus groups at various locations around the United States. The primary reason for conducting focus groups on a national basis is to begin to understand the extent of the sampling frame that will be required if the contingent valuation study of total values is carried out. While exact locations are yet to be worked out we would propose a minimum of two focus groups in each of the following areas: Arizona, California, southeastern United States, and somewhere in the heavily populated Boston to Washington, DC, corridor.

The results of focus groups must be carefully interpreted. Because focus groups are typically conducted with a small number of participants, it is impossible to extrapolate the findings to the general population from which the participants were drawn. Besides the problem of making population inferences based on small samples, it is also doubtful that the participants are a good representation of the population from which they were drawn. Participation in a focus group discussion requires a relatively large commitment of time, the ability to travel to a specified meeting site at a specific time, and the willingness to engage in discussions with strangers. Because of these factors, we often find that only 10 to 20 percent of eligible participants agree to participate in the focus group. As a consequence, it is very difficult to ensure that the participants are representative of the target population at large.

For these reasons, we view focus groups as an important step toward quantitative research for assessing attitudes toward and values for possible impacts to changes in dam operations and other alternatives, not as a substitute for quantitative research. Focus groups are a fertile source of hypotheses that can then be more systematically, quantitatively tested. Based on results from the focus groups, we plan to design and execute an attitude survey to more systematically explore these hypotheses.

At the end of the qualitative research effort, a report will be prepared including information on resource service flows presented to participants, focus group agendas, a summary of the results from each focus group, a draft attitude survey, an assessment of the potential for a successful attitude survey, and a proposed revised research plan for remaining parts of the total

valuation study if continuation appears warranted. This will form the basis for decisions on the part of GCES with cooperating agencies about whether to proceed and under what conditions.

The Attitude Survey

So much of the attitude survey design will depend on results from the qualitative research that it is hard to be specific. To a considerable extent, the attitude survey will involve further refining, quantifying, and generalizing to relevant populations what was learned in the focus groups.

As we now conceive of it, the first section of the attitude survey would focus on respondents' experience with and knowledge about the Grand Canyon. This would allow us to determine how respondent experience affects the attributes that they feel are part of the system and how important these attributes are to more and less experienced respondents.

The second section would present lists of various attributes which may or may not be affected by flows and would obtain importance ratings for each attribute. Various holistic scenarios of environmental change that might be related to dam operations would be presented to determine if such scenarios are understandable to the general public. Based on the evaluations of these attributes we would be able to identify those most appropriate for contingent valuation scenarios.

The attitude survey would also be used to evaluate a variety of proposed payment vehicles for use in the contingent valuation survey. Respondents would be asked about the appropriateness and fairness of the vehicles, as well as other potential sources of bias.

Finally, the attitude survey would assess a variety of social and economic characteristics of the respondents to determine whether some attributes are more important and salient for some types of groups and which attributes are salient for all groups.

Analysis of the attitude survey would determine the most important attributes and examine the effects of experience, regional, and socioeconomic characteristics on importance ratings. Judgments about appropriateness of payment vehicles and informational needs for the contingent valuation survey would also be made based on the attitude survey. In addition, the attitude survey will provide important information as to whether a mail survey is a viable alternative for carrying out the contingent total valuation study. A sample of perhaps as many as two thousand would be necessary for the attitude survey, especially if the focus groups indicate that regional differences are considered to be important.

Upon completion of the attitude survey and analysis of the results, a report will be prepared that will describe procedures, summarize the how people view the resources in question, assess regional and other differences among subsamples, and make recommendations regarding the contingent valuation survey. At this point, officials associated with GCES, including its cooperating agencies, will decide whether to continue with the research and under what conditions.

Contingent Valuation Survey

Results of the attitude survey will then be used to design the contingent valuation survey. Unless the attitude survey indicates otherwise, a mail survey will be conducted. A sample (or samples) will be selected from the appropriate population as determined by the focus group and attitude survey results. Whether one sample or more than one is drawn would to some extent be dictated by the number of resources and the number of alternatives that must be studied. Given that large numbers of resources are potential candidates for inclusion in the total valuation questionnaire, written descriptions may become so long and tedious that all the potential changes in resource service flows cannot be evaluated within a single survey form. Additional samples might have to be used to cover the full range of impacts. The size of the samples would be dictated by the degree of precision required for GCES and whether regionally valid estimates are needed.

As noted previously, changes in resource service flows will be measured relative to service flows under current operations, which will serve as the baseline. Thus, the valuation portion of the questionnaire will begin with a description of current operations and future prospects for the resources that have been chosen for inclusion in the study based on the attitude survey results. An alternative will be introduced in terms of the implications it would have for these resources compared to the effects of current operations. These scenarios comparing the alternative to the baseline would be as complete as possible and include both beneficial and detrimental effects. Quantitative information would be presented to the maximum extent possible. For example, statements like, "Some beaches will be lost," are not likely to be adequate. Rather, more detailed, quantitative descriptions of the resources and the effects of the baseline and alternatives would be presented.

Once the status and future conditions of the resources have been compared between the baseline and the alternative, the contingent valuation question would elicit willingness to pay for the alternative. Theoretically, the questions will be rooted in the total valuation framework developed in Section 1, above. Dichotomous choice contingent valuation questions will be used. Consideration will be given to using double-bounded dichotomous choice questions. While normal dichotomous choice questions simply ask respondents to accept or reject a specific dollar amount, the double-bounded approach asks a second question where the amount to be considered is higher for those who accepted the first offer and lower for those who rejected the first offer. This approach has been shown to have desirable statistical properties, but responses are more difficult to analyze and results may be biased by contamination of responses to the second question by the offer amount in the first question. The nature of this potential bias is comparable to starting-point bias often encountered in contingent valuation bidding games. Additional alternatives would be evaluated up to the point where respondent fatigue begins to affect responses as determined from the pretest and pilot studies discussed below. The survey would include not only contingent valuation questions but also detailed questions on socioeconomic characteristics and other personal attributes which would help us understand responses to the contingent valuation questions.

As part of the survey design process, the survey instrument or instruments will be formally pretested. The primary purpose of this pretest is to assure that all parts of the survey are working well, but it would also have two other goals. Offer amounts for the dichotomous

choice questions would be determined based partly on the pretest. Also, we would work with small samples until we are reasonably confident that respondents can deal with the number of contingent valuation questions in the semi-final instrument or instruments. Next, these semi-final instruments will be used in pilot studies with larger samples, about 200 respondents, drawn from the national sampling frame unless a more restricted sampling frame has been chosen. This would be the final check on the survey and the offer amounts before the survey is administered to the full samples.

Results would be derived using established statistical procedures. The analysis will produce not only final value estimates, but also link the valuation responses to socioeconomic, attitudinal, and, for reasons discussed in Section 1 above, motivational variables.

The final report for the total value study will summarize procedures at each stage in the project, results of the focus groups and attitude surveys, and valuation results expanded from the samples or samples to the population. If it proves practically and technically feasible, a valuation surface (as described earlier) will be estimated so that additional alternatives beyond those specifically considered in the EIS can be considered later.

Conclusions

As developed here, this project has the potential to be more than a valuation study. The total value results might even be considered as mere "frosting on the cake." The "cake" itself will be based on the focus groups and the attitude survey and would provide a thorough understanding of how people view the resources at the bottom of the Grand Canyon in the broader context of Grand Canyon National Park, Glen Canyon National Recreation Area, and other resource management units as a whole. Which resources within the total set of resources studied under GCES are of greatest and least interest to the public will have been identified. Undergirding the values themselves will be data on why people hold those values. Specific issues such as potential non-use values for electricity generated at the dam will have been addressed. Our understanding of the extent to which scientific uncertainty affected the conclusions will have been enhanced.

Can this study be performed with sufficient precision that it will meet high scientific standards? We conclude that it can. Section 1 of this report showed that there is a theoretical framework sufficient to form a foundation for this study. Literature surveyed in Section 2 showed that often heard concerns about the validity of contingent valuation are exaggerated. For use values, the suggestion of Cummings, et al. (1986), that, even under the best of circumstances, accuracy only up to ± 50 percent is possible seems to us more like a worst case scenario than an estimate of precision based on the data. Admittedly, there is less evidence to use in evaluating the potential precision of contingent total values. Based on the research on contingent existence values discussed in Section 2, one could expect accuracy to ± 50 percent. We suspect that this level of confidence will prove to be overly pessimistic as more research accumulates. Given the potential magnitudes involved, accuracy at this level is sufficient to make the results potentially useful in policy analysis. This is particularly true given all that would be learned, over and above the dollar values, about how the general public views the resource issues being addressed under GCES. To neglect total values in favor of more narrowly defined use values would, we believe, leave a major gap in the economic studies under GCES, a gap that would carry over into the Glen Canyon Dam EIS.

REFERENCES CITED

Beers, D., J.R. Bieke, E.K. Madsen, M.J. Brennan, and F.P. Prager. 1989. "Comments of ASARCO, Inc. and Newmont Mining Corporation." Presented before the United States Department of The Interior, November.

Bishop, R.C., K.J. Boyle, M.P. Welsh, R.M. Baumgartner, and P.R. Rathbun. 1987. Glen Canyon Dam Releases and Downstream Recreation: An Analysis of User Preferences and Economic Values. January.

Bishop, R. and T.A. Heberlein. 1984. "Contingent Valuation Methods and Ecosystem Damages from Acid Rain." Department of Agricultural Economics, Staff Paper No. 217, University of Wisconsin -Madison.

Bishop, R., T.A. Heberlein, and M.J. Kealy. 1983. "Contingent Valuation of Environmental Assets: Comparison With A Simulated Market." Natural Resources Journal, Volume 23, pp. 619-633.

Bishop, R.C., and T.A. Heberlein. 1990. "The Contingent Valuation Method," in R.L. Johnson and G.V. Johnson, eds., Economic Valuation of Natural Resources. p. 81.

Bishop, R.C. 1986. "Resource Valuation Under Uncertainty: Theoretical Principles for Empirical Research." Advances in Applied Microeconomics, V. Kerry Smith, ed. Greenwich, Connecticut: JAI Press, pp. 133-152.

Bishop, R.C., and M.P. Welsh. 1990. "Existence Values in Benefit-Cost Analysis and Damage Assessment." Land Economics. December.

Bishop, R.C. 1982. Option Value: An Exposition and Extension. Land Economics 58:1-15.

Bishop, R.C., and T.A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics, Volume 61 (2), pp. 376-379, May.

Bishop, R.C., T.A. Heberlein, D.W. McCollum, and M.P. Welsh. 1988. A Validation Experiment for Valuation Techniques. Working Paper, Center for Resource Policy Studies and Programs, College of Agricultural and Life Sciences, University of Wisconsin-Madison, (also, Electric Power Research Institute, Palo Alto, CA).

Bowker, J. and J.R. Stoll. 1988. "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." In American Journal of Agricultural Economics, Volume 70, Number 2, pp. 372-381, May.

Boyce, R.R., T.C. Brown, G.D. McClelland, G.L. Peterson, and W.D. Schulze. 1989. "Experimental Evidence of Existence Value in Payment and Compensation Contexts," Benefits and Costs in Natural Resources Planning, Interim Report No. 2, Western Regional Research Project W-133. pp. 305-336.

Boyle, K. and R.C. Bishop. 1987. "Valuing Wildlife in Benefit Cost Analyses: A Case Study Involving Endangered Species." Water Resources Research, Volume 23, Number 5, May.

Brookshire, D.S., D.L. Coursey, and K.M. Radosevich. 1988. Market methods and assessment of benefits: Some further results. In: G.L. Peterson, et al., eds., Amenity Resource Valuation: Integrating Economics with Other Disciplines, pp. 167-178.

Brookshire, D.S., L.S. Eubanks, and C.S. Sorg. 1986. "Existence Vaues and Normative Economics: Implications for Valuing Water Resources." Water Resources Research 22:1509-1518.

Brookshire, D.S., L.S. Eubanks, and A. Randall. 1983. "Estimating Option Prices and Existence Values for Wildlife Resources." Land Economics, Volume 59, Number 1, pp. 1-15, February.

Brookshire, D.S., B.C. Ives, and W.D. Schulze. 1976. "The Valuation of Aesthetic Preferences." Water Resources Research. 22:1509-1518.

Brookshire, D.S., M.A. Thayer, W.D. Schulze, and R.C. d'Arge. 1982. "Valuing public goods: A comparison of survey and hedonic approaches. American Economic Review. 72:165-177.

Brown, Jr., M. Gardner, and M.L. Plummer. 1989. Methods for Valuing Acid Deposition and Air Pollution Effects. National Acid Precipitation Assessment Program, State of the Science/Technology Report 27, Part A, Washington D.C.: U.S. Protection Agency.

Byerlee, D.R. 1971. "Option Demand and Consumer Surplus: Comment." Quarterly Journal of Economics. 85:523-27.

Carson, R.T., W.M. Hanemann, and R.C. Mitchell. 1986. "Determining the Demand for Public Goods by Simulating Referendums at Different Tax Prices." Manuscript, University of California, San Diego.

Chavas, J., R.C. Bishop and Kathleen Sergerson. 1986. Ex-Ante Consumer Welfare Evaluation in Cost-Benefit Analysis. Journal of Environmental Economics and Management 13:255-268.

Cicchetti, C.J. and N. Peck. 1989. "Assessing Natural Resource Damages: The Case Against Contingent Value Survey Methods." Natural Resources and Environment, Volume 4.

Cicchetti, C.J., and A.M. Freeman III. 1971. "Option Demand and Consumer Surplus: Further Comment." Quarterly Journal of Economics. 85:528-539.

Cory, D.C., and B.C. Saliba. 1987. "Requiem for Option Value." Land Economics. 63:1-10.

Coursey, D.L., J.L. Hovis, and W.D. Schulze. 1987. "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value." Quarterly Journal of Economics. 102:679-690.

Cummings, R., D.S. Brookshire and W.D. Schulze. Valuing Environmental Goods: An Assessment of the Contingent Valuation Method, Rowman and Allanheld, publishers, 1986.

d'Arge, R. and J.F. Shogren. 1989. Okoboji Experiment: Comparing Non-Market Valuation Techniques In An Unusually Well-Defined Market for Water Quality. Ecological Economics, Volume 1, pp. 251-259.

Desvousges, W.H, and V.A. Skahen. 1987. "Techniques to Measure Damages to Natural Resources." U.S. Department of the Interior, CERCLA 301 Task Force. June.

Desvousges, W.H., V.K. Smith, and M.P. McGivney. 1983. "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvement." Report to the U.S. Environmental Protection Agency, EPA-230-05-83-001, Washington, March.

Dickie, M., A. Fisher, and S.Gerking. 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study." Journal of the American Statistical Association, Volume 82, no. 397, pp. 69-75.

Fischhoff, B. and L. Furby. 1988. "Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility." Journal of Risk and Uncertainity, Volume 1, pp. 147-184.

Fisher, A. and R. Raucher. 1984. "Intrinsic Benefits of Improved Water Quality: Conceptual and Empirical Perspectives." Advances in Applied Micro-Economics, Volume 3, pp. 37-66.

Freeman, A.M. 1990. "Nonuse Values in Natural Resource Damage Assessment." Bowdoin College and Resources for the Future, Inc., January.

Freeman, A.M. 1984. "The Sign and Size of Option Value." Land Economics, Volume 60, Number 1, February.

Freeman, A.M. 1986. "Uncertainty and Environmental Policy: The Role of Option and Quasi-Option Values. Advances in Applied Macro-Economics, V. Kerry Smith, ed., Volume 4, pp. 153-157, JAI Press, Inc., Greenwich, Connecticut.

Graham, D.A. 1981. "Cost-Benefit Analysis Under Uncertainty," American Economic Review, vol. 71, pp. 715-725.

Hageman, R. 1985. Valuing Marine Mammal Populations: Benefit Valuations In A Multi Species Ecosystem. Administrative Report No. LJ-85-22, National Marine Fisheries Service, Southwest Fisheries Center, La Jolla, California, September.

Harrison, D. and J.A. Hausman. 1989. Advanced Notice of Proposed Rulemaking Regarding Revision of Natural Resource Damage Assessment Regulations. Paper presented before the Department of Interior, November.

Hartman, R., and M.A. Plummer. 1987. Option Value Under Income and Price Uncertainity. Journal of Environmental Economics and Management 14:212-225.

Heberlein, T.A., and R.C. Bishop. 1986. "Assessing the Validity of Contingent Valuation: Three Field Experiments." Science of the Total Environment, Volume 56, pp. 99-107.

Hoehn, J.P., and A. Randall. 1985. "Demand Based and Contingent Valuation: An Empirical Comparison." Selected paper annual meeting of American Agricultural Economics Association, Ames, Iowa, August.

Hoehn, J.P., and A. Randall. 1987. "A Satisfactory Benefit-Cost Indicator from Contingent Valuation." Journal of Environmental Economic Management, Volume 14, pp. 226-247.

Hoehn, J.P., and A. Randall. 1989. "Too Many Proposals Pass the Benefit Cost Test." American Economic Review. 789(3):544-551.

Kahneman, D. 1986. "Comments on the Contingent Valuation Method." In R.G. Cummings, D.S. Brookshire, and W.D. Schulze, eds., Valuing Environmental Goods, Rowman and Allanheld, Totawa, New Jersey.

Kahneman, D. and J.L. Knetsch. Forthcoming. "Valuing Public Goods: The Purchase of Moral Satisfaction."

Kealy, M.J., J.F. Dovidio, and M.L. Rockel. 1987. "Willingness to Pay to Prevent Additional Damages to the Adirondacks From Acid Rain," Regional Science Review, Volume 15, pp. 118-141.

Kealy, M.J., J.F. Dovidio, and M.L. Rockel. 1988. "Accuracy in Valuation Is a Matter of Degree," Land Economics, 64 (No. 2): 158-171.

King, D.A., D.J. Flynn, and W.W. Shaw. 1988. "Total and Existence Values of a Herd of Desert Bighorn Sheep," Western Regional Research Publication W-133, Benefits and Costs in Natural Resources Planning, Interim Report, pp. 243-264.

Knetsch, J.L., and R.K. Davis. 1966. "Comparison of Methods for Recreation Evaluation," pp.125-142. A.V. Kneese and S.C. Smith, eds., Water Research. The Johns Hopkins University Press, Baltimore.

Krutilla, J.V. 1967. "Conservation Reconsidered." American Economic Review, Volume 57, pp. 777-786.

Krutilla, J.V. and A.C. Fisher. 1975. "The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources." John Hopkins University Press, Baltimore, Maryland.

Lindsay, C.M. 1969. "Option Demand and Consumer Surplus." Quarterly Journal of Economics, Volume 83, pp. 344-345, May.

Long, M.F. 1967. "Collective Consumption Servicer of Individual Consumption Goods." Comment. Quarterly Journal of Economics, Volume 81, pp. 351-52.

Loomis, J. 1989. "Test-Retest Reliability of the Contingent Valuation Method" A Comparison of General Population and Visitor Responses." American Journal of Agricultural Economics, Volume 71, Number 1, February.

Loomis, J. 1987. "Balancing Public Trust Resources of Mono Lake and Los Angeles' Water Right: An Economic Approach." Water Resources Research, Volume 23, No. 8, pp. 1449-1456, August.

Madriaga, B. and K.E. McConnell. 1987. "Exploring Existence Value." Water Resources Research, Volume 23, Number 5, pp. 936-942, May.

McConnell, K.E. 1983. "Existence Value and Bequest Value". R.D. Rowe and L.G. Chestnut eds., Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas. Westivew Press, Boulder, Colorado.

Mitchell, R.C., and R.T. Carson. 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method.

Mendelsohn, R. 1984. "Estimating the Structural Equations of Implicit Markets and Household Production Functions." Review of Economics and Statistics, Volume 66, pp. 673-677.

Phillips, C. and R.J. Zeckhauser. 1989. "Contingent Valuation of Damage to Natural Resources: How Accurate? How Appropriate?" Toxics Law Reporter, pp. 520-522, Bureau of Natural Affairs, publisher, October.

Randall, A. and J.R. Stoll. 1983. "Existence Value In A Total Valuation Framework." Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas, Robert D. Rowe and Lauraine C. Chestnut, eds., Westview Press, 1983.

Randall, A., and W. Kriesel. 1990. "Evaluating National Policy Proposals by Contingent Valuation," Rebecca L. Johnson and Gary V. Johnson, eds., Economic Valuation of Natural Resources: Issues, Theory, and Applications. pp. 153-176.

Ready, Richard C. 1990. "Welfare Measures Under Uncertainty," unpublished paper, Department of Agricultural Economics, University of Kentucky, Lexington, Kentucky.

Samples, Karl C., Marcia M. Gowen, and John A. Dixon. 1986. "The Validity of the Contingent Valuation Method for Estimating Non-Use Components of Preservation Values for Unique Natural Resources," Unpublished paper, Department of Agricultural and Resource Economics, University of Hawaii.

Schmalensee, R. 1972. "Option Demand and Consumer's Surplus: Valuing Price Changes Under Uncertainity." American Economic Review, Volume 62, pp. 813-24.

Schulze, W.D., D.S. Brookshire, E.G. Walther, K.K. MacFarland, M.A. Thayer, R.L. Whitworth, S. Ben-David, W. Malm, and J. Molenar. 1983. "The Economic Benefits of Preserving Visibility in the National Parklands of the Southwest." Natural Resources Journal, Volume 23, pp. 149-173, January.

Sellar, C., J.R. Stoll, and J. Chavas. 1985. "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques. Land Economics, Volume 61, pp. 156-175.

Smith, V.K. 1987a. "Nonuse Values In Benefit Cost Analysis." Southern Economics, Volume 54, pp. 19-26.

Smith, V.K. 1987b. "Uncertainity, Benefit-Cost Analysis, and the Treatment of Option Value." Journal of Environmental Economics Management, Volume 14, pp. 283-292.

Stoll, J. and L.A. Johnson. 1985. "Concepts of Value, Nonmarket Valuation, and the Case of the Whooping Crane." Transactions of the North American Wildlife and Natural Resources Conference, pp. 382-393.

Sutherland, R.J., and R.G. Walsh. 1985. "Effect of Distance on the Preservation Value of Water Quality." Land Economics, Volume 61, Number 3, pp. 281-291.

Thayer, M.A. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence." Journal of Environmental Economic Management, Volume 8, pp. 27-44.

United States Department of the Interior. 1988. Glen Canyon Environmental Studies Final Report, January.

Walsh, R.G., L.D. Sanders, and J.B. Loomis. 1985. Wild and Scenic River Economics: Recreation Use and Preservation Values, Report to the American Wilderness Alliance (Department of Agriculture and Natural Resource Economics, Colorado State University).

Walsh, R., J.B. Loomis, and R.A. Gillman. 1984. "Valuing Option, Existence, and Bequest Demands for Wilderness." Land Economics, Volume 60, Number 1, February.

Weisbrod, B.A. "Collective-Consumption Services of Individual Consumption Goods." Quarterly Journal of Economics, Volume 78, pp. 471-477, 1964.

Welle, P. Forthcoming. "Option Values for Acid Deposition Control: Some Empirical Evidence."

Welsh, M.P. 1986. "Exploring the Accuracy of the Contingent Valuation Method: Comparisons with Simulated Markets" (Ph.D. Dissertation, University of Wisconsin, Madison).

Whitehead, J.C. and G.C. Blomquist. 1990. "Substitute Environmental Goods and Existence Value." Department of Economics. East Carolina University, and University of Kentucky.

APPENDIX A

REVIEWERS' COMMENTS

April 22, 1991

THE ROLE OF NON-USE VALUES IN THE GLEN CANYON ENVIRONMENTAL STUDIES

Summary Review Comments by Ann Fisher on
Assessment of the Potential for Non-Use Valuation Research Under the Glen Canyon Environmental Studies: Literature Review and a Study Prospectus,
March 14, 1991 Draft Report

At the March 22-23 meeting, the reviewers were asked to summarize their judgments about three aspects of this proposal:

1. Does the concept of non-use values apply to operation of the Glen Canyon Dam?

2. Is there a need to estimate non-use values for alternative operating scenarios? Will significant information be gained that can help clarify choices among the proposed alternatives?

3. Can non-use values for the Glen Canyon Dam operations be estimated with enough scientific rigor to withstand the scrutiny of the National Academy of Sciences, the public, and other agencies involved with the operation of the dam?

The reasoning behind my answers to these questions is described below. It has been shaped by my experience with the U.S. Environmental Protection Agency and as an academic economist.

Non-use Values as Part of Total Values

Section 1 of the draft report reviews the literature and confirms that non-use values fit comfortably within conceptual models for examining how total values change when a policy action affects individuals' utility or well-being.

Various authors define the components of total value differently, causing some fuzziness in the boundaries between use values and non-use values. For example, an argument can be made that bequest value is part of use value, where the use occurs through one's heirs. Others lump bequest value with existence value, which often is defined as the value of having or preserving a resource for its own sake, regardless of whether anyone ever will use it. Similar arguments are made about option value, which accounts for both the possibility that people might want to use the resource in the future (demand uncertainty) and concerns about whether it will be available for use in the future (supply uncertainty). [Bequest values are part of existence values in the draft report, and option values are treated separately when

uncertainty is introduced in the model.]

Use values are easier to measure than non-use values. Use values already measured for the Glen Canyon Dam include recreation values for anglers and rafters, and consumption of electricity by those who rely on its hydropower. Because non-use values are more difficult to measure, some analysts are skeptical about whether they belong in policy analysis. This issue can be clarified by looking at what benefit-cost analysis can and cannot do, as well as how it has been used in the past.

The Role of Benefit-Cost Analysis in Policy Decisions

Policy alternatives almost always vary along many dimensions. These dimensions include many economic benefits and costs, legal restrictions, and political and equity concerns. Decision makers responsible for selecting among alternative policies typically receive information and recommendations from their analysts before making a choice. Even after long deliberation and agonizing over the analysts' recommendations, the decision maker's choice often seems to be made in a big "black box." Reasons for selecting a particular alternative often are far from clear, and provide neither guidance for future decisions nor justification for the one that was made.

Economists developed benefit-cost analysis to make this multi-dimensional decision process more manageable and accountable. They reduced the number of dimensions that must be juggled by using market data to estimate all the economic benefits and costs of each policy alternative. Then they recommended the one with the highest net benefits because it would allocate resources most efficiently--and became frustrated when decision makers seemed to ignore their input. Let's examine why, and whether this is inevitable.[1]

Traditional benefit-cost analysis has relied on the assumption that all benefits and costs can be identified, quantified, and monetized. But the reality facing decision makers typically falls into one of the following categories:

1. All benefits and costs can be identified and quantified, but economists have not been able to assign values to some of them or the values assigned are controversial.

2. All benefits and costs can be identified, but only

---

[1]Much of the discussion in this section is drawn from my "Comment" in Natural Resource Economics: Policy Problems and Contemporary Economic Analysis, ed. by D.W. Bromley, Boston: Kluwer-Nijhoff Publishing, 1986, pp. 201-209.

some can be quantified. Of those, only some can be monetized.

3. Because of genuine uncertainties about what effects the policy action might have, some benefits or costs simply cannot be identified.

Even when the conditions for a traditional benefit-cost analysis are met, its results shed light only on the **efficiency** of various alternatives. The decision maker then must weigh the equity, legal, and political aspects and determine whether they are sufficiently important to overturn the decision that would be most efficient.

Many early analyses simply ignored any benefits and costs that were difficult to monetize. This was especially common for benefits or costs that could not be quantified or identified. Thus while benefit-cost analyses (BCAs) illuminated some of the efficiency considerations, they failed to indicate the likely importance of any benefits or costs that were not monetized, quantified, or identifiable. They also failed to acknowledge the potential importance of non-efficiency considerations. Analysts presumed that efficiency was the only valid decision criterion, rather than making the case that BCA can reduce the many efficiency dimensions and allow the decision maker to concentrate on the relative importance of other dimensions that are valid considerations in the decision but difficult to measure.

Some recent BCAs have had a significant role in policy decisions, for example in environmental regulations.[2] This role can continue and be expanded by defining BCA more broadly. Analysts can maximize the usefulness of benefit-cost analyses by:

a. Identifying, quantifying, and monetizing as many benefits and costs as possible, and providing net benefit estimates for each policy alternative.

b. Quantifying as many as possible of the remaining benefits and costs, and listing those benefits and costs that can be identified but not quantified. The analyst should specify how large or important they would have to be to overturn the ranking based on the BCA.

c. Indicating whether there are likely to be benefits or costs that cannot be identified at the time the analysis is conducted. If so, the analyst might speculate about whether they are likely to be important

---

[2] EPA's Use of Benefit-Cost Analysis: 1981-1986, Office of Policy, Planning and Evaluation, U.S. Environmental Protection Agency, Washington, DC, EPA-230-05-87-028, August 1987.

enough to overturn the ranking based on (a) and (b). Of course, the speculative nature of such statements should be made clear.

Following these steps yields messy results compared with a traditional benefit-cost analysis. However, this process reduces the even messier set of complex dimensions originally facing the decision maker. It can illuminate substantial portions of the black box, so that the non-efficiency considerations receive more careful scrutiny in the ultimate policy choice--even if only because the efficiency costs of satisfying other decision criteria become clear. The BCA makes it easier to justify decisions and to use them as guidance for future decisions. At the same time, it provides ample latitude for the policy maker to do what he or she is paid for: to make decisions on the basis of relevant factors.

BCA can provide important insights for evaluating the proposed alternatives for operating the Glen Canyon Dam. The alternatives are too complex to provide all the identification, quantification, and monetization that would be required for a traditional BCA (category (1) above). In fact, this situation most closely matches category (3) above. This means steps (b) and (c) must supplement the more traditional step (a) to maximize the usefulness of the BCA conducted for the Glen Canyon Dam operating alternatives.

The next question examines the role of non-use values in analyzing how benefits change across alternative dam operating scenarios.

Does the Non-Use Value Concept Apply to Changes in Glen Canyon Dam Operations?

Substantial progress has been made in developing inputs for a BCA of alternative Glen Canyon Dam operating practices. Plans for the Environmental Impact Statement (EIS) include providing cost estimates for 10 operating alternatives. Use benefits have been estimated for rafters and anglers, although the proposed operating alternatives may not match those examined for the recreation use estimates. Use benefits also have been estimated for consumers relying on electricity from the dam. At least two other groups could have substantial use benefits: hikers and campers.

Adding estimates of use benefits for hikers and campers will not satisfy the conceptual definition of changes in total benefits associated with the proposed alternatives. This is because both users and non-users could have non-use values affected by changes in dam operations. Examples could include values for Native American archeological sites, endangered species such as the humpback chub, and the general ecosystem in its present state (e.g., with the resurgence of the southern bald eagle and the peregrine falcon).

Changes in total value are composed of changes in both use values and non-use values. The fuzziness in boundaries for these value categories could mean that it is easier to estimate changes in total values, rather than trying to aggregate separate estimates of changes in use value and changes in non-use value. (This covers both a conceptual issue and an empirical issue. Conceptually, the division between use values and non-use values is not especially distinct. Empirically, efforts to measure only use values could end up including some non-use components, and efforts to measure only non-use values could end up including some use values. Adding these separate estimates then would include double counting. This can be avoided by estimating the overall change in value.)

Will Research on Non-Use Values for the Glen Canyon Area Provide Significant Information for Policy Decisions?

The magnitude of non-use values for the Glen Canyon Area is an empirical question. Even small non-use values held by many people can sum to amounts large enough to be significant in a benefit-cost analysis. Thus part of the study is to determine how extensively non-use values are held. The study prospectus describes a task to identify whether the sample can be restricted to the southwestern states or needs to cover a larger geographic area, perhaps as large at the nation.[3]

If non-use values for the Glen Canyon Area are small, then it will be more difficult to estimate how they would change across alternative dam operating practices, just because it is harder to estimate changes in small magnitudes than changes in big magnitudes. Estimates of changes in non-use values also will be influenced by how different the alternative dam operating scenarios are. Alternatives with similar impacts on non-use resources make it harder to get estimates that distinguish among them. The researchers can control the second influence, by describing alternatives that are quite different (although they might be more extreme than the alternatives in the EIS). But even for quite different alternatives, existence values might not show up if they truly are very small.

The few credible studies of existence values for other resources suggest that they can be substantial. The only way to

---

[3]Conceptually, foreigners also could have existence values for the Glen Canyon Area that would be affected by the dam's operation. Measuring those values would be a practical problem, and it might be possible to argue that they are smaller than the values held by those ultimately sampled. If not, an arbitrary decision can be made to limit the scope to U.S. residents because the decision is being made by federal agencies.

find out how large existence values are for proposed changes in the Glen Canyon Dam's operations is to measure them. A relatively inexpensive pilot study (as proposed for the next stage at the March 22-23 meeting) will let decision makers know whether non-use values are large for a small number of people, small for a large number of people, or so small that they get lost in the noise. Even determining that they are very small would be helpful in assessing the alternative dam operating scenarios, because changes in costs then become more important in the decision.

Can Non-Use Values Be Measured with Enough Scientific Rigor to Withstand Scrutiny of Relevant Parties?

We still are in the pioneering stages of applying contingent valuation (CV) for non-use values. Given the potential importance of non-use values and the fact that no other measurement option has been developed, the question is whether CV's imperfect information is better than none.

The draft report lists many problems with conducting contingent valuation studies of non-use values. These include the fact that some impacts will be hard to identify and harder to describe to lay people, and the difficulty of getting values for changes in the Glen Canyon area rather than for the entire Grand Canyon. Contingent valuation still has not been embraced by all economists, although it has gained credibility for measuring use values because of the consistency among its results when compared with alternative methods for estimating the same use values.

One basis for evaluating CV is to examine the standards applied to scientific estimates of what ecosystem changes will occur for the proposed alternatives. One aspect of this is reflected in the examination of the Reference Operating Conditions (ROCs), as discussed in the draft. Another important aspect is the expected size of an estimate's errors. If CV's expected errors are smaller or approximately the same as those for estimates of the ecosystem effects, then the CV estimates are as helpful as the predictions of physical changes. Scientists predicting changes in ecosystems often are pleased if their error bands are within an order of magnitude. The (admittedly sparse) evidence for CV suggests that its estimates will have much narrower error bands. Both types of estimates are likely to distinguish better between proposed policies that are quite different than among alternatives that are quite similar.

So long as the limitations of a pioneering methodology accompany estimates and their use, the consensus seems to be that contingent valuation results can be useful for policy decisions. Several examples can be mentioned.

The Draft Summary cites the Water Resources Council's

Principles and Guidelines, which permit CV estimates.[4] EPA recently relied on contingent valuation in its Regulatory Impact Analysis (RIA) to estimate the benefits of controlling power plant impacts on visibility in the Grand Canyon.[5] A nationwide sample of households gave their willingness-to-pay for improved visibility at the Grand Canyon. Respondents allocated over two thirds of their estimates to bequest and existence values, or what is termed non-use value for our purposes. Because the estimates in the RIA interpret and transform the existing economic research, the authors caution that the reported results "should be interpreted as depicting the direction, character, and expected order of magnitude of some of the economic benefits rather than depicting the exact amounts,"(p.29) Even so, estimates likely to understate values suggest per household willingness to pay of $1.30 to $2.50 per year to achieve visibility improvements for the 90 percent sulfate control level. When both use and non-use values are included, all of the options analyzed (from 50 percent to 90 percent control) showed positive net benefit estimates.

EPA similarly planned to use CV in the RIA for regulating surface coal mining.[6] However, both EPA and the Department of Interior potentially had the authority to regulate, and DOI ended up with the responsibility for this regulation.

Australia has collected contingent valuation data to estimate the preservation benefits from protecting the Kakadu National Park and a related area from potential damage associated with mining.[7] It is not yet clear how the information will be used in their decision process.

The most important action affecting the potential for using CV could be the July 1989 Court of Appeals decision that non-use values are a valid component for natural resource damage assessments. The potential for large non-use value estimates may

---

[4]U.S. Water Resources Council, Economic and Environmental Principles and Guidelines for Water and Related Land Resource Implementation Studies, March 1983.

[5]Regulatory Impact Analysis of a Revision of the Federal Implementation Plan for the State of Arizona to Include SO2 Controls for the Navajo Generating Station, February 5, 1990 draft, including January 30, 1991 addendum, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.

[6]Personal communication, Allen Basala, Office of Air Quality Planning and Standards, U.S. EPA, March, 1991.

[7]Personal communication, Leanne Wilks, Australia Resource Assessment Commission, January 1991.

have been a major factor in the March 1991 settlement of the Alaska oil spill natural resource damage claims, and in settlements for mining damage in Colorado streams and acid rain damage in Minnesota.

Poorly conducted contingent valuation studies will not and should not withstand scrutiny of knowledgeable professionals, but well-done CV studies have withstood such scrutiny. Substantial progress has been made during the past decade in applying this technique. For the Glen Canyon Dam, CV estimates to measure the change in total values will provide a check on the use value estimates already made, as well as providing new information on non-use values. This has several advantages (described earlier) over conducting a CV study that attempts to isolate non-use values. Of course, some notion of the relative magnitudes of use values and non-use values can be obtained by asking people whether they are present users, expect to be users in the future, or never expect to use the Glen Canyon Dam area.

Contingent valuation studies require enough resources to support the primary data collection and the analysis. They are not inexpensive relative to other methods employed to measure use values. Properly conducted CV studies, however, can shed substantial light on the size of total values, including non-use values. This information can be especially helpful when there are substantial differences in costs across alternatives. A relatively modest investment in making sure the estimates of total values are reasonably complete can insure that costly, low-benefit options are avoided in the final policy decision.

The other reviewers have chosen different emphases in their comments. The bottom line is consensus in our judgments about the potential for using CV to estimate changes in total values associated with Glen Canyon Dam operating alternatives. If the study prospectus is revised in response to the comments expressed during the March 22-23, 1991 workshop, the study should yield results useful for analyzing the proposed alternatives as well as advance the state of the art in CV.

April 3, 1991

Comments on
Draft Report and Proposal

by

A. Myrick Freeman III
Bowdoin College

We were asked to respond to three sets of questions about the desirability and feasibility of estimating nonuse values attributable to changes in the operation of the Glen Canyon Dam as part of the Glen Canyon Dam environmental impact statement process. The questions and my answers to them are the basis of these comments.

I. ARE NONUSE VALUES IMPORTANT IN THIS CASE?

I would answer this in the affirmative; but I would qualify it by saying that this is true in principle. But we cannot be sure that nonuse values will be large in this situation without making some effort to determine what attributes of the Grand Canyon floor are valued by people and what the magnitudes of these values are. As the draft report makes clear, the concept of nonuse values is plausible. And nonuse values can be fitted into the standard theory of economic value in a consistent and coherent way. We also have evidence that nonuse values can be large in magnitude, at least for changes in the status of endangered or rare species and for significant changes in the key attributes of sites such as water quality for lakes and rivers or visibility, for example, at the Grand Canyon.

However, we not know very much yet about which services of a resource are most valued by nonusers. The early conceptual and empirical literature focused on such things as the preservation of unique geological structures and ecological systems or preservation of the endangered species. But in the case at hand, the existence of the Canyon floor is not at stake. Rather, there may be changes in a whole vector of service flows from the resource. We have not yet had much experience in attempting to measure nonuse values in this type of case.

Most of the earlier literature on nonuse values has focused on deviations from a base line of the pristine, preintervention level of environmental services. This baseline had normative significance, at least where environmental stewardship motivations were involved. But here, what vector of environmental services should be chosen as a baseline is not clear. And for any possible choice, it is conceivable that some individuals would value changes from that baseline positively while other value them negatively. This possibility should be acknowledged and accounted for in the study design.

Finally, we do not know whether nonuse values stem from a flow of environmental services over time or from the knowledge of preservation as distinct from a destruction or an irreversible loss. In stark form, the question is whether individuals will hold significant nonuse values for a policy for which only postpones a sure (or at least highly likely) loss for, say, five years. The study design should provide for some examination of

the dynamic and intertemporal patterns of environmental change and alternative ways of modeling these in a utility theoretic framework. These issues have implications both for how resource changes are described in a contingent valuation study and the form of contingent payments.

II. WHAT CAN BE ACHIEVED BY THE PROPOSED STUDY?

The first stage of the proposed study, the conceptual and qualitative research will help to provide answers to the questions raised above: What environmental services are value by nonusers? What is the relevant baseline? How does the temporal pattern of change in resource services affect nonuse values? These are important questions in the case of the Grand Canyon. And they are important questions for other kinds of environmental and resource valuation problems, as well. The information gained from this stage of the study will have value not only to users of the study (those who are preparing the EIS) but to many economists and policy analysts involved in similar issues.

The pilot study stage will also provide an opportunity to test important hypotheses about the sources of nonuse values from this resource and about the most appropriate way of framing questions to elicit these values.

Finally, if the results of the first two stages of the proposed research are encouraging and a decision is made to proceed to a full-fledged survey, then we should have quantitative information on the nature and magnitude of changes

in nonuse values stemming from changes in the Glen Canyon Dam operating regime.

III. CAN ESTIMATES OF NONUSE VALUES BE DONE WITH SCIENTIFIC RIGOR?

First, what is meant by the term "scientific rigor"? One possible answer would be based on replicatability. But I do not think this is a very useful way of defining scientific rigor in this case. A CV study would be replicated if another analyst used the same survey instrument, used the same procedures and sampling frame to pick respondents, and use the same statistical techniques to answer the data, and the final results were equivalent within some predetermined statistical bounds. But this kind of replication can only establish that if you ask the question to similar people you get the same or similar answers. It does not establish that the answers contain relevant, useful economic information.

A second interpretation of the term would involve comparison of the values obtained from the CV instrument with the "true" value. But as the Draft Report makes clear, for this kind of problem the true value can never be known. So this kind of test is simply not applicable.

In my view, the test of scientific rigor has to proceed along the following lines. First, it should be noted that practioners in the CV field have identified a number of problem areas and issues which can affect both the validity and reliability of CV responses. For example there is general

agreement that the form of the valuation question is important. Closed ended or referendum questions work better than the bidding game format. A number of sources of potential bias have been identified. These include implied valued cues and part-whole bias. One important aspect of judging the scientific rigor of a CV study is to have the best experts in the field make a judgment as to whether these problem areas and issues have been addressed and dealt with in the best way possible, according to today's state of knowledge and expertise.

Also, of course, an aspect of scientific rigor is the sampling design and sample size. Another aspect is the use of appropriate statistical techniques and model in analyzing the data. These relatively straightforward issues to deal with in evaluating CVM studies.

In answering a question about the possible scientific rigor of nonuse value measures, let me observe first that many of the early studies of nonuse values cited in the Draft Report do not pass the test of scientific rigor by today's standards. One reason for this is that the state of the art has advanced substantially in recent years; and today we would apply a much more strict set of standards in judging these studies.

Also, there is not presently a well established and accepted set of survey instruments which can simply be taken off the shelf to measure nonuse values. In this respect, the measurement of nonuse values by CV is different from the measurement of use

value for recreation, for example. This is a problem which is at the frontier of the development of this measurement technique.

But on the basis of the Draft Report, the discussion at the workshop in Phoenix, and the past work of the principal investigators, I am confident that the research team has the potential for producing estimates of nonuse value which will pass the test of scientific rigor that I have described here. I use the term "potential" advisedly, since the conduct of a scientifically defensible CV study would require adequate time and financial resources. It would also be very valuable to have the continued advice and regular review of research designs and progress by outside experts in the field.

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

COLLEGE OF NATURAL RESOURCES
AGRICULTURAL EXPERIMENT STATION
GIANNINI FOUNDATION OF AGRICULTURAL ECONOMICS
DEPARTMENT OF AGRICULTURAL & RESOURCE ECONOMICS

207 GIANNINI HALL
BERKELEY, CALIFORNIA 94720

March 21, 1991

Dr. Michael Welsh
HBRS Inc
585 Science Drive, Suite A
Madison, WI 53711.

Dear Mike,

As I mentioned in our phone conversation, I will unfortunately be unable to attend the meeting in Phoenix due to last minute personal commitments in Berkeley. I enclose the comments on the Draft Report that I had intended to make at the meeting.

I must say that this is a very solid study plan. It is careful and well thought out. There is an overwhelming case for including in GCES an economic study of non-use values. Without such a study, any EIS or policy analysis relying on the GCES work will be seriously flawed and open to challenge. The plan of work mapped out on pages 61 - 72 represents a cautious, systematic, credible approach to estimating relevant non-use values, and I strongly endorse it.

Sincerely,

Michael

Michael Hanemann
Associate Professor

COMMENTS

Alan Randall
The Ohio State University

The Draft Report on non-use valuation for GCES, prepared by Richard C. Bishop and Michael P. Welsh, is a very sound document. It identifies the important issues. The analysis is technically sound, and approaches the state of the art. Current controversies are addressed and judicious conclusions are reached. For a document of its kind, it is unusually well-written.

My comments start with the big questions: Are non-use values relevant, in general? Are they likely to be important to GCES? Can they be estimated with scientific rigor. Having answered these questions in the affirmative, I then move to a number of more technical points.

1. It is difficult to make the argument that benefits and costs should always be decisive, in policy decisions about natural and environmental resources. Other perspectives on political philosophy have their place, and these other perspectives set limits on the consideration that should be accorded to benefits and costs.

Nevertheless, benefits and costs count, under a surprisingly wide range of philosophical perspectives. Benefits and costs count because satisfaction of human preferences counts. Philosophers may argue about what kinds of concerns trump human

preferences under what conditions, but it is difficult to imagine a serious philosophy in which human preferences count for nothing at all.

The justification for taking benefits and costs seriously is that benefits and costs are acceptable expressions of human preferences. It follows that whatever people have preferences about should be included in an accounting of benefits and costs. Not whatever people buy in organized markets; not whatever people can be observed using; but whatever people have preferences about.

That non-use values count in benefit cost analysis follows directly from the argument that benefits and costs count in policy decisions. Combine this logic with the empirical hypothesis that some people have strong preferences about the environment downstream from Glen Canyon dam, independent of use -- i.e., that non-use values are substantial -- and the implication is unambiguous. Any study of the benefits and costs of alternative management strategies should include non-use values.

2. Non-use values can be included in a total value framework for benefit cost analysis, in a scientifically rigorous fashion. At the theoretical level, the total value framework is rigorous, and it provides rigorous specification of total value and of the various component values including non-use value.

Scientific rigor requires both conceptual and empirical rigor, and there is less consensus about the empirical rigor of non-use value estimates. I believe that some of the controversy is rooted in an inappropriate concept of empirical rigor. Let me explain.

Kahneman and Knetsch, among others, argue that a measuring tool that gives different values for the same thing is unacceptable (see Draft Report, p. 43). Following this line of thought, any evidence that estimates of non-use values vary with changes in the design of the contingent valuation (CVM) devices for measuring them is damning to non-use values and to CVM. And the literature reports plenty of such evidence. I suggest a very different perspective.

Market prices and asset values are conditional. That is, they depend on institutions, supply and demand conditions, and expectations about both. Those skeptical about regarding market prices as informative about "more fundamental" values often raise the issue of price volatility. Prices seem to fluctuate "too much"; they are too hard to predict; they are prone to "speculative bubbles," etc. Was the notorious Van Gogh original really worth the less-than-$10 million that appraisers thought in the early 1980s, the more-than-$50 million that Mr. Alan Bond paid for it in the late 1980s, or the less-than-$30 million for which he sold it a couple of years later? Recently, economists have converged on an answer to that vexing question: "all of the above." That is, the rational markets hypothesis posits that the market price at any moment in time reflects all of the information available at that moment. Each of the prices for the Van Gogh was right for its time, given all that was known at that time. Price is conditional.

Observed prices come in time series: a series of prices, each the result of the "natural experiment" that generated the conditions unique to its moment in time. But that moment will never be repeated, so the natural experiment can never be exactly replicated. The fact

that we observe (under ideal conditions) a single price at a given moment must not obscure the principle that market price is conditional.

Contingent values are like market values in this respect. They, too, are conditional. Contingent valuation provides an opportunity to induce cross-sectional variation in WTP (WTA), as a purposeful experimental strategy.

Viewed in this light, the sensitivity of CVM results to variations in the valuation conditions provides no evidence for questioning the validity of non-use values and/or CVM. It does, however, emphasize the importance of determining the appropriate valuation conditions to design into the CVM instruments. This task is two-part.

a. A program of research is needed, to map the relationship between valuation conditions and reported WTP (or WTA). The focus groups suggested in the Draft Report will be helpful. Focus groups are but one component of a full-fledged program of qualitative research. I would recommend that qualitative research not be restricted to focus groups, but be conceived broadly to include one-on-one dialogues, individual administration of draft questionnaires, intensive de-briefing, etc.

Following qualitative research, small sample experiments with alternative CVM instruments would permit empirical mapping of the relationships between valuation conditions and reported WTP (WTA). This process, done carefully, is empirically rigorous.

b. If it happens that WTP (WTA) is empirically sensitive to variations within the plausible range of valuation conditions, then an appropriate set of valuation conditions must be chosen for implementation in the final CVM study. Once the conditionality of contingent values is fully understood, it becomes apparent that some conscious decisions must be made

as to what are the "right" conditions for valuation. In this context, "right" means appropriate for the policy purposes at hand.

To re-iterate, the sensitivity of CVM results to valuation conditions is a fact of life, as is the sensitivity of market values to market conditions. It is no cause to question the scientific rigor of non-use value estimation. Given the conditionality of values, a program of rigorous research can be designed to map the relationships between valuation conditions and WTP (WTA).

3. The appropriate conceptual basis for a study of non-use values is the total value framework. It is now well-known that attempts to estimate total value by adding-up the values of its various components expose the effort to a variety of conceptual and empirical errors. I do not belief the Draft Report authors and I have any disagreements on this point. Nevertheless, I encourage them to remain firm against any pressures to treat non-use values as separate from other kinds of value, and non-use value studies as independent of other valuation efforts.

4. There is much to be gained by designing CVM studies to estimate a value surface, i.e., an (n+1)-dimensional relationship between WTP and the quantities of the n relevant categories of environmental services. If the environmental services are appropriately defined, values will be lower-bounded at zero in each dimension of services.

Provided an estimated value surface, benefits of particular policy alternatives can be calculated, given estimates of the productivity of policies in terms of environmental services.

The alternative approach, treating each policy alternative as a CVM scenario, seems both more difficult technically and less likely to produce generalizable results.

5. Embedding -- the conjecture that reported WTP for a particular environmental service varies depending upon what else is to be (explicitly or implicitly) provided along with it -- is not necessarily an artifact of CVM. There is both theoretical and empirical evidence that embedding is a real-world phenomenon. That, of course, leaves open the question of whether CVM is some way distorts (perhaps exacerbates) real-world embedding. It is possible that poorly specified CVM scenarios may produce exaggerated embedding. The most highly-touted embedding results were obtained with very poorly specified CVM scenarios. However, as yet there is little evidence that well-designed CVM studies generate any special embedding problems. In this, I agree with the authors of the draft report.

6. Is it possible that nonuse values might be positive for introduced species (e.g., trout), electricity, and even the idea of "progress" in the southwest (which might be encouraged by managing releases from Glen Canyon dam so as to minimize the price of electricity)? In principle, it is. There is no conceptual basis for limiting (say) existence values to things natural and pristine. Many environmental economists operate on the premise that the largest existence values are associated with things natural, pristine, and relatively scarce. This is perhaps a reasonable intuition, but it must remain an empirical hypothesis until thoroughly tested.

It follows that the qualitative research phase of GCES non-use value studies should thoroughly explore the possibility of positive non-use values for trout, electricity, "progress", etc. In economics, generally, we don't assume that everything of value can be found on one side of the equation. Instead, we seek to identify the alternative with the largest net value. The same should go for non-use values.

7. The issue of national versus local sampling frames cannot be resolved entirely until a thorough program of qualitative research has been completed. There have been suggestions that a national sample might generate excessively large WTP. However, I think this fear is exaggerated.

a. If the alternative scenarios are carefully specified (e.g., baseline and alternative scenarios may differ as to riverine conditions but are identical in terms of conditions beyond the valley floors), there should be little worry that respondent feelings about the Grand Canyon will contaminate CVM responses concerning riverine conditions.

b. Further, recent CVM studies seem to refute the hypothesis the respondents will report WTP of $5 or $10 for anything "good", whether they really care about it or not. (If this kind of response behavior were prevalent, national studies of rather modest local environmental improvements would generate huge aggregate WTP).

Qualitative research at the national level could determine rather inexpensively whether these kinds of problems might occur, and whether modifications to CVM design might solve the problems. It is also possible, of course, that very few individuals beyond the regional population and visitors to the area have positive non-use WTP for these

environmental services. If so, it would not be cost-effective to invest large amounts in national surveying. Again, some qualitative research and nationwide pre-testing could resolve this issue.

8. The issue is raised (Draft Report, p. 67) as to whether utility bills is an appropriate vehicle for expressing WTP. In general, the preferred vehicle is no vehicle at all. It is best to specify payment as a reduction in disposable income. Something like "the costs of the program will increase (taxes and) the prices of things you buy, so that households like yours will have $_____ less to spend each year" comes close to communicating the idea of reduced disposable income. For nonuse values in a national survey, it is difficult to see why something like that would not work. For a regional survey, perhaps one could try "the proposed program would increase utility costs and reduce economic activity in the region, so households like yours would have $ _____ less to spend each."

May 9, 1991

Comments on the Assessment of the Potential for Nonuse Valuation Research Under the Glen Canyon Environmental Studies

V. Kerry Smith*

The objectives of the Bishop-Welsh report are:

- to describe the state of the theory underlying nonuse values for environmental resources;
- to review and evaluate the studies that attempted to measure one or more components of nonuse values;
- to assess the feasibility of measuring nonuse values for the particular issues posed by alternative management plans for the Glen Canyon Dam; and
- to propose a strategy for estimating these values, provided the assessment indicated this would be warranted based on theory, past practice, and the specific problem.

I have divided my comments on their report into two parts. In the first, I offer some reactions to their summary and analysis of this literature. The second part responds to three questions to focus on the relevance and feasibility of a study of nonuse values as part of the Glen Canyon Environment studies. They are:

(1) Do nonuse values apply to the resource management issues associated with the Glen Canyon Dam?

(2) Does effective management of this resource require measures of nonuse values?

(3) If nonuse values are needed, can they be measured with the same scientific rigor as use values and the other information contributing to our understanding of the management issues?

I. Reactions to the Summary

Bishop and Welsh provide an excellent overview of the nonuse value literature and identify a sound conceptual basis for nonuse values within a Hicksian framework for applied welfare analysis. Their treatment clearly distinguishes option and existence values and the roles for each in defining the value of an environmental resource. Obviously they have an excellent command of this literature and a broad perspective on its historical development. Three aspects of their appraisal of the conceptual foundations are especially noteworthy. First, they argue (correctly in my judgment) that total value, reflecting all motives for valuing the services provided by a resource, should be the focus of any attempt to incorporate nonuse values in evaluations of alternative management plans for the Glen Canyon Dam.

Second, their description of the resources involved acknowledges that the management issues raise questions associated with changing the mix of "services" provided by the resources below the dam.[1] These services will be involve different resources, all within an interrelated system. It is acknowledged that the dam has transformed the downstream ecosystem and a steady-state system may not be present today. Nonetheless, management activities will yield further changes in the levels and quality of the available resource services. This orientation provides a need to refocus much of the existing conceptual literature from considering one resource contributing to people's well-being to considering multiple resources. This will require

evaluating changes from some specified baseline condition to alternative endpoints rather than in relation to levels where no services would be available (i.e., existence of a resource or species).

Finally, they acknowledge that there are two important aspects of valuing these resources: measuring total value for the "representative" individual (and how it changes with their characteristics and the particular mix of services described) and determining how many people hold these values--or, in other words, gauging the extent of the market for the resource.

I agree with their overall verdict that there is a well-established conceptual basis for measuring nonuse values in a total value framework. Nonetheless, I do feel that further conceptual work is warranted as part of the process of developing total value estimates. Four areas may deserve further attention.

1. Use/Nonuse Connections

For the most part, existing literature has focused on separating use and nonuse values. As a result, past analyses have used the point of separation (i.e., values for the resource at the choke price) as a central element in definitions of nonuse values. Bishop and Welsh adopt McConnell's [1983] position that nonuse values arise from what should be recognized as public good services provided by certain types of natural assets. A logical next step in the process is to consider the connections between use-related values and nonuse values rather than the separations. More specifically, we might envision environmental resources as providing different types of services, some supporting recreational activities (e.g., fishing, whitewater rafting, etc.) and others associated with nonuse benefits. When both are assumed to be arguments in an

individual's preference function, then we can expect the marginal rate of substitution associated with the use-related services (in comparison to a numeraire) to be influenced by the public, nonuse-related services. Of course, there are restrictions to preferences that would preclude some connections, and weak complementarity is one example.

Larson's [1990] recent proposal for measuring existence value is an example of how these types of postulated connections can be used with observed behavior to impute values. In his application, they completely rely on the maintained connections. New conceptual work should address whether there is some middle ground between assuming away nonuse values with weak complementarity and directly specifying their contributions with Larson's proposal.

2. Defining the Services Underlying Nonuse Values

As Bishop and Welsh acknowledged (and I noted earlier), the issues raised by evaluating management alternatives for the Glen Canyon Dam involve changes in the mix of services provided by the resources influenced by its operations. This is different from conventional, single-resource descriptions of the role of nonuse-related services in people's preference functions.

Several important differences may warrant further attention. The natural system will impose some constraints on how these services are interrelated and, given these connections, the feasible combinations of services. The findings of the "research flows" currently underway on the river offer one way of learning more about these connections. They need to be reflected in both scenario design for the valuation (CVM) survey and in the explanations presented to respondents.

Equally important, the conceptions of substitutes for the resource involved need to be explored when evaluating the system as an interrelated and evolving ecosystem. Do these interconnections make the resource unique, or should we evaluate the values (use and nonuse) from newly created recreational and habitat resources by considering the substitutes for each component as a separate entity?

3. "Negative" Nonuse Values

Because the management options will usually entail changes in the level (and/or quality) of the services underlying nonuse values rather than questions of some services' existence, some people may require compensation to be indifferent toward one operating condition that enhances some species' habitats at the expense of other resources.

This possible outcome implies the need to separate those respondents favoring and those opposing changes. To assume the latter have zero values for the change can be incorrect. This requires that the analysis consider the prospect of designing a valuation survey that must elicit both willingness-to-pay and willingness-to-accept responses. Because of questions relating to past empirical findings (see Mitchell and Carson [1989] or Knetsch [1990] for summaries), recent theoretical developments (Hanemann [1991]), and potential implementation issues, the prospect of "negative" nonuse values for some people may need to be explored in focus groups and the qualitative aspects of designing data collection instruments.

4. Extent of the Market

Bishop and Welsh clearly identify this question as central to their survey design. It may also be desirable to use the theoretical and empirical literature describing how these issues are addressed for marketed commodities, as well as the recently revised work of Bockstael and McConnell [1991] evaluating the properties of the choke price as the quality of a nonmarketed good changes.

The report provides a good review of the empirical efforts to measure nonuse value and what has been learned from them. In the process, the authors clearly indicate that they have evaluated the relative quality (and therefore the plausibility) of some studies relative to others. More specific information seems desirable on whether consistencies were observed in the measures of nonuse value and effects of income, sociodemographic variables, connections to use values, etc. Such a summary might suggest features to evaluate preliminary surveys testing the research design for the present project, as well as an analysis of the final data.[2]

II. Questions About a Nonuse Value Study as Part of CGES

1. Do Nonuse Values Apply?

The answer here is a decisive "yes." Moreover, there is broad professional consensus among resource economists on this point. Within the economics community generally, this concept has realized a growing level of acceptance. Indeed, as part of a commentary on research in public finance and modeling altruistic behavior (in the 1990 AEA meetings), Professor Kenneth Arrow, Nobel laureate from Stanford University, used the empirical

estimates for existence values as examples of how the features of preferences associated with altruism and nonuse values are potentially important parts of people's motivations.

In a judicial setting, the D. C. Court of Appeals ruling on the Department of Interior's regulations for natural resource damage assessments (in Type B assessments) held that nonuse values were a legitimate component of the values provided by a natural resource. DOI's recently proposed revisions to their regulations (in response to the Court's ruling) clarify their position, acknowledging that nonuse values should be included as a component of a natural resource's values and, therefore, have been defined as part of what they describe as "compensable value."[3] Of course, this does not mean they will be markedly different from zero for all people (or, indeed, positive for all people in all situations).

2. Is There a Need for Nonuse Values in GCES?

Again the answer is a clear-cut "yes." Indeed, even among analysts who question the overall importance of nonuse values, all agree that they are likely to be important for unique, widely recognized natural resources (and the Grand Canyon is routinely cited as just such a case). Nonuse values may be especially important for some of the management alternatives being considered for the Glen Canyon. A small number of recreationists may gain reasonably large individual benefits from some changes, but the aggregate value over all individuals realizing use and nonuse values could easily be overwhelmed by the reduced nonuse values that could accompany the same changes for a larger group of individuals who would not use the resource.

This is clearly worth knowing, and the answers are not obvious--as evidenced by the differences in the existence values for visibility measured for the Grand Canyon with apparently

different perceptions of damage to other aspects of resources in two different studies (see Chestnut and Rowe [1990] and Balson et al. [1990]).

3. Can Nonuse Values Be Estimated Adhering to Rigorous Scientific Standards?

The answer here is simply not clear. It is impossible to know until we do more research. However, committing to initiate the study does not require it's completion. A phased program with specific decision points should allow analysts to determine whether completing the study will be feasible before the full costs associated with the survey activities (the most expensive part of the research) must be incurred. This is important, because if the decision is made to fully implement the study, the scientific plausibility of the findings could be affected by a failure to provide adequate resources to support the survey tasks required.

It is important to initiate the work. Even if the initial phases suggest that nonuse values cannot be reliably measured with the available budget, it seems likely that information about the resources with the most significant nonuse values can be identified and more insight on the extent of the market for determining aggregate use-related benefits will be capable of being evaluated.

Notes

*University Distinguished Professor, Resource and Environmental Economics Program, Department of Economics, N. C. State University, and Resource for the Future University Fellow.

1. I am using the term "services" to include all the ways in which a resource can contribute to the well-being of people, including those conventionally associated with use values as well as any nonuse values.

2. One early study with a detailed analysis of multivariate functions describing existence values is Bennett's [1984] analysis. This should be added to their references.

3. However, the most recent DOI rules do raise questions with the reliability of CVM for measuring nonuse values. They designate it the least reliable method, but also acknowledge that no other are methods available.

Paul R. Portney
Vice President and Senior Fellow

11 January 1992

Mr. Thomas A. Campbell
General Counsel
National Oceanic and
Atmospheric Administration
U.S. Department of Commerce
Room 5816
14th and Constitution Avenue, NW
Washington, DC 20230

Dear Tom:

In behalf of co-chairs Kenneth Arrow and Robert Solow, as well as Edward Leamer, Roy Radner, and Howard Schuman, I am very pleased to submit the enclosed final report of the Contingent Valuation Panel (Panel), appointed by the Office of General Counsel to consider the reliability of the contingent valuation (or CV) methodology in measuring passive-use values of natural resources.

The Panel concludes that CV studies can produce estimates reliable enough to be the starting point for a judicial or administrative determination of natural resource damages -- including lost passive-use value. To be acceptable for this purpose, though, such studies should adhere closely to the guidelines described in the report. It is not necessary that every injunction be completely obeyed; however, the more closely the guidelines are followed, the more reliable the result. A CV study that is carefully constructed, administered, and analyzed will contain information that judges, juries and other decisionmakers will wish to use, in combination with other evidence, including the testimony of expert witnesses.

The report is organized in the following way. Section I is the Introduction to the report. In Section II, the drawbacks of the CV technique are discussed. Section III discusses several key issues concerning the use of the CV technique. Section IV includes guidelines to which the Panel believes any CV study should adhere if the study is to produce information useful in natural resource damage assessments. In Section V a research agenda is described. Section VI presents the Panel's conclusions in more detail.

1616 P Street, NW • Washington, DC 20036 • Telephone 202-328-5000 • Fax 202-939-3460 • Cable Resource

In transmitting this report, we wish to thank you, Randall Luthi, Linda Burlington and the other members of NOAA's Damage Assessment Regulations Team for the outstanding support and assistance provided to the Panel. Your collective efforts were above and beyond the call of duty and our report is the better for them.

We hope the report will be useful in your rulemaking efforts.

Sincerely,

Paul R. Portney

Enclosure

Report of the NOAA Panel on Contingent Valuation

January 11, 1993

Kenneth Arrow
Robert Solow
Paul R. Portney
Edward E. Leamer
Roy Radner
Howard Schuman

Report of the NOAA Panel on Contingent Valuation

Date: January 12, 1993

I. INTRODUCTION

Under the Oil Pollution Act of 1990, the President--acting through the Under Secretary of Commerce for Oceans and Atmosphere--is required to issue regulations establishing procedures for assessing damages to or destruction of natural resources resulting from a discharge of oil covered by the Act. These procedures are to ensure the recovery of restoration costs as well as the diminution in value of the affected resources and any reasonable costs of conducting the damage assessment.

At least some of the values that might be diminished by such a discharge are relatively straightforward to measure through information revealed in market transactions. For instance, if the discharge kills fish and thereby reduces the incomes of commercial fishermen, their losses can reasonably be calculated by the reduced catch multiplied by the market price(s) of the fish (less, of course, any costs they would have incurred). Similarly, if the discharge of oil discourages tourist travel to an area, the lost incomes of those owning and/or operating motels, cottages, or other facilities can be reasonably represented by the difference in revenues between the affected period and a "normal" season. Even the losses to recreational fishermen, boaters, swimmers, hikers, and others who make active use of the areas affected by the discharge can be included in the estimate of diminished value, although these losses will

generally be somewhat more difficult to value than the more obvious out-of-pocket losses.

The losses described above have come to be known as lost "use values" because they are experienced by those who, in a variety of different ways, make active use of the resources adversely affected by the discharge. But for at least the last twenty-five years, economists have recognized the possibility that individuals who make no active use of a particular beach, river, bay, or other such natural resource might, nevertheless, derive satisfaction from its mere existence, even if they never intend to make active use of it.

This concept has come to be known as "existence value" and it is the major element of what are now referred to as "non-use" or "passive-use" values (the latter term is employed in the balance of this report). In regulations promulgated by the Department of the Interior in 1986 under the Comprehensive Environmental Response, Compensation, and Liability Act -- regulations that also pertained to natural resource damage assessments -- passive-use values were included among the losses for which trustees could recover. The inclusion of passive-use values was recently upheld by the D. C. Court of Appeals (State of Ohio v. Department of the Interior, 880 F.2d 432 (D.C. Cir. 1989)), as long as they could be reliably measured.

This begs an interesting and important question, however. If passive-use values are to be included among the compensable losses for which trustees can make recovery under the Oil

Pollution Act, how will they be estimated? Unlike losses to commercial fishermen or recreational property owners, there are no direct market transactions that can be observed to provide information on which estimates can be based. Unlike losses to boaters, swimmers, recreational fishermen and others, there exist no indirect methods through which market data can provide at least some clues as to lost values. In other words, there appear to be neither obvious nor even subtle behavioral trails that can provide information about lost passive-use values.

Some experts believe that there exists an approach that can provide useful information about the economic significance of the lost passive-use values individuals may suffer when oil discharges damage natural resources. Known as the contingent valuation (or CV) technique, this approach is based on the direct elicitation of these values from individuals through the use of carefully designed and administered sample surveys. Its appeal lies in its potential to inform damage assessment in an area (lost passive-use values) where there appear to be no behavioral trails to be followed.

Typically, CV studies provide respondents with information about a hypothetical government program that would reduce the likelihood of a <u>future</u> adverse environmental event such as an oil spill, chemical accident, or the like. Respondents are usually given some specific information about the exact nature of the damages that the program in question would prevent. And they are also confronted in the study with a question or questions that

provide information about the economic sacrifice they would have to make to support the environmental program. This may take the form of an open-ended question asking what is the maximum amount they would be willing to pay for the program in question; it may involve a series of questions confronting them with different prices for the program depending on their previous answers; or it may take the form of a hypothetical referendum (like a school bond issue) in which respondents are told how much each would have to pay if the measure passed and are then asked to cast a simple "yes" or "no" vote. (The conceptually correct measure of lost passive-use value for environmental damage that has already occurred is the minimum amount of compensation that each affected individual would be willing to accept. Nevertheless, because of concern that respondents would give unrealistically high answers to such questions, virtually all previous CV studies have described scenarios in which respondents are asked to pay to prevent future occurrences of similar accidents. This is the conservative choice because willingness to accept compensation should exceed willingness to pay, if only trivially; we say more about other biases below.)

The CV technique has been used for twenty years or so to estimate passive-use values. In the last five years, however, there has been a dramatic increase in the number of academic papers and presentations related to the CV technique. This is due in part to the availability of comprehensive reference texts on the subject (Mitchell and Carson (1989), for instance), and to

the growing interest both nationally and internationally in environmental problems and policies. But it is also attributable to the growing use of the CV technique in estimating lost passive-use values in litigation arising from state and federal statutes designed to protect natural resources. Since Ohio v. Department of the Interior admitted the concept of passive-use values in damage assessments, this can only give added impetus to the use of CV in such litigation.

The CV technique is the subject of great controversy. Its detractors argue that respondents give answers that are inconsistent with the tenets of rational choice, that these respondents do not understand what it is they are being asked to value (and, thus, that stated values reflect more than that which they are being asked to value), that respondents fail to take CV questions seriously because the results of the surveys are not binding, and raise other objections as well. Proponents of the CV technique acknowledge that its early (and even some current) applications suffered from many of the problems critics have noted, but believe that more recent and comprehensive studies have already or soon will be able to deal with these objections.

This (sometimes acrimonious) debate has put the National Oceanic and Atmospheric Administration (NOAA) in a very difficult spot. NOAA must decide in promulgating the regulations under the Oil Pollution Act whether the CV technique is capable of providing reliable information about lost existence or other passive-use values. Toward this end, NOAA appointed the

Contingent Valuation Panel to consider this question and make recommendations to it.

This report is the product of the Panel's deliberations and is organized in the following way. Following this introduction, the drawbacks to the CV technique are discussed in Section II. Section III discusses several key issues concerning the design of CV surveys, including use of the referendum format to elicit individual values, ways of addressing the so-called "embedding" problem, and the evaluation of damages that last for some period but not forever. Section IV presents guidelines to which the Panel believes any CV study should adhere if the study is to produce information useful in natural resource damage assessment. (These are elaborated upon in an Appendix.) In Section V a research agenda is described; it is the Panel's belief that future applications of the CV technique may be less time-consuming and contentious if the research described in the agenda is carried out. Section VI presents the Panel's conclusions.

II. CRITICISMS OF THE CONTINGENT VALUATION METHOD

The contingent valuation method has been criticized for many reasons and the Panel believes that a number of these criticisms are particularly compelling. Before identifying and discussing these problems, however, it is worth pointing out that they all take on added importance in light of the impossibility of validating externally the results of CV studies. It should be

noticed, however, that this same disadvantage must inhere in any method of assessing damages from deprivation of passive-use. It is not special to the CV approach although, as suggested in Section I, there are currently no other methods capable of providing information on these values.

One way to evade this difficulty, at least partially, is to construct experiments in which an artificial opportunity is created to pay for environmental goods. The goods in question can perfectly well involve passive use. Then the results of a CV estimate of willingness to pay can be compared with the "real" results when the opportunity is made available to the same sample or an analogous sample.

A few such experiments have been attempted. The most recent, due to Seip and Strand (1992), used CV to estimate willingness to pay for membership in a Norwegian organization devoted to environmental affairs, and compared this estimate with actual responses when a number of the same respondents were presented with an opportunity actually to contribute. The finding was that self-reported willingness to pay was significantly greater than "actual" willingness to pay. A recent study by Duffield and Patterson (1991) took as the environmental amenity in question the maintenance of stream flow in two Montana rivers. The rivers in question provided spawning grounds for two rare species of fish; passive use was believed to be the main motivation for respondents. One of two parallel samples was asked about hypothetical willingness to contribute to the Montana

Nature Conservancy which would then maintain stream flow; the other was offered an opportunity actually to contribute to the same organization for the same purpose. It was found that response rates and expressed willingness to contribute were significantly higher when the contribution was hypothetical than when "expressed willingness" meant an immediate cash contribution. On the other hand, the size of contributions, hypothetical in one case and actual in the other, was not much different as between those who said they would contribute and those who did so.

These studies suggest that the CV technique is likely to overstate "real" willingness to pay. Duffield and Patterson, however, hold out hope that the differences are small enough and predictable enough that CV estimates could be discounted for possible overstatement and then used as a conservative estimate of willingness to pay. Clearly more such experiments would be useful.

A less direct test of the "reality" of CV estimates of lost passive use values is to use the technique to estimate willingness to pay for ordinary market goods and then to compare the results with actual purchases. This has been tried by Dickie, Fisher, and Gerking (1987) using the demand for strawberries. When the data were re-analyzed by Diamond, Hausman, Leonard, and Denning (1992), it was found that the CV approach tended systematically to overestimate quantity demanded at each price, sometimes by as much as 50 percent. This result

has to be qualified in two ways. First, the original CV study seems to have been fairly casual by the standards now proposed by practitioners; pre-testing and improvement of the survey instrument might (perhaps) have narrowed the gap. And second, it seems to go too far to conclude from systematic over-estimation that the CV study, even as conducted, provides no information about the demand for strawberries. Much of the same could be said about a study submitted to the Panel by Cummings and Harrison (1992) comparing hypothetical and demonstrated willingness to pay for small household goods. (See also Bishop and Heberlein (1979).)

External validation of the CV method remains an important issue. A critically important contribution could come from experiments in which state-of-the-art CV studies are employed in contexts where they can in fact be compared with "real" behavioral willingness to pay for goods that can actually be bought and sold.

Of the other problems arising in CV studies, the following are of most concern to the Panel: (i) the contingent valuation method can produce results that appear to be inconsistent with assumptions of rational choice; (ii) responses to CV surveys sometimes seem implausibly large in view of the many programs for which individuals might be asked to contribute and the existence of both public and private goods that might be substitutes for the resource(s) in question; (iii) relatively few previous applications of the CV method have reminded respondents

forcefully of the budget constraints under which all must operate; (iv) it is difficult in CV surveys to provide adequate information to respondents about the policy or program for which values are being elicited and to be sure they have absorbed and accepted this information as the basis for their responses; (v) in generating aggregate estimates using the CV technique, it is sometimes difficult determining the "extent of the market;" and (vi) respondents in CV surveys may actually be expressing feelings about public spiritedness or the "warm glow" of giving, rather than actual willingness to pay for the program in question. We discuss each of these briefly.

Inconsistency with Rational Choice

Some of the empirical results produced by CV studies have been alleged to be inconsistent with the assumptions of rational choice. This raises two questions: What requirements are imposed by rationality? Why are they relevant to the evaluation of the reliability of the CV method?

Rationality in its weakest form requires certain kinds of consistency among choices made by individuals. For instance, if an individual chooses some purchases at a given set of prices and income, then if some prices fall and there are no other changes, the goods that the individual would now buy would make him or her better off. Similarly, we would expect an individual's preferences over public goods (i.e., bridges, highways, air quality) to reflect the same kind of consistency.

Common notions of rationality impose other requirements which are relevant in different contexts. Usually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual. Also, if marginal or incremental willingness to pay for additional amounts does decline with the amount already available, it is usually not reasonable to assume that it declines very abruptly.

This point assumes importance in view of some empirical evidence from CV studies that willingness to pay does not increase with the good. In one study, Kahneman (1986) found that willingness to pay for the cleanup of all lakes in Ontario was only slightly more than willingness to pay for cleaning up lakes in just one region. Evidence of this kind has multiplied (see Kahneman and Knetch (1992), Desvousges, et al. (1992), and Diamond et al. (1992)). Desvousges' result is very striking; the average willingness to pay to take measures to prevent 2,000 migratory birds (not endangered species) from dying in oil-filled ponds was as great as that for preventing 20,000 or 200,000 birds from dying. Diminishing marginal willingness to pay for additional protection could be expected to result in some drop. But a drop to zero, especially when the willingness to pay for the first 2,000 birds is certainly not trivial, is hard to explain as the expression of a consistent, rational set of

choices.

It has been argued on a more technical level that the studies finding such apparent inconsistencies are defective, that the choices are not presented clearly to the respondents. In the study referred to immediately above, for instance, respondents were told that 2,000 birds was "...much less than 1%" of the total migratory bird population while 200,000 birds was "...about 2%" of the total. This may have led respondents to evaluate the programs as being essentially the same. But on the face of it, the evidence certainly raises some serious questions about the rationality of the responses.

It could be asked whether rationality is indeed needed. Why not take the values found as given? There are two answers. One is that we do not know yet how to reason about values without some assumption of rationality, if indeed it is possible at all. Rationality requirements impose a constraint on the possible values, without which damage judgments would be arbitrary. A second answer is that, as discussed above, it is difficult to find objective counterparts to verify the values obtained in response to questionnaires. Therefore, some form of internal consistency is the least we would need to feel some confidence that the verbal answers corresponded to some reality.

Implausibility of Responses

The CV method is generally used to elicit values for a specific program to prevent environmental damage, whether it be

dead animals, spoilage of a pristine wilderness area, or loss of visibility in some very unusually clear area. Though in each case, individuals often express zero willingness to pay, average willingness to pay over the whole sample is often at least a few dollars and frequently $20 to $50. With 100,000,000 households in the United States, these responses result in very large totals, frequently over $1 billion. Some have argued that these large sums are in themselves incredible and cast doubt on the CV method. The Panel is not convinced by this argument, since it is hard to have an intuition as to a reasonable total.

But there is a different problem with these answers. One can envision many possible types of environmental damage -- oil spills or groundwater contamination in many different locations, visibility impairment in a variety of places, and so on. Would the average individual or household <u>really</u> be willing to pay $50 or even $5 to prevent each one? This seems very unlikely, since the total resulting willingness to pay for all such programs could easily become a very large fraction of one's income or perhaps even exceed it.

In other words, even if the willingness to pay responses to individual environmental insults are correct if only one program is to be considered, they may give overestimates when there are expected to be a large number of environmental problems. Similarly, if individuals fail to consider seriously the public or private goods that might be substitutes for the resources in question, their responses to questions in a CV survey may be

unrealistically large.

Absence of a Meaningful Budget Constraint

Even if respondents in CV surveys take seriously the hypothetical referendum (or other type of) questions being asked them, they may respond without thinking carefully about how much disposable income they have available to allocate to all causes, public and private (see Kemp and Maxwell (1992), for instance). Specifically, respondents might reveal a willingness to pay of, say, $100 for a project that would reduce the risk of an oil spill; but if asked what current or planned expenditures they would forgo to pay for the program, they might instead re-evaluate their responses and revise them downward. This is similar to the problem identified immediately above where individuals fail to think of the possible multiplicity of environmental projects or policies they might be asked to support. To date, relatively few CV surveys have reminded respondents convincingly of the very real economic constraints within which spending decisions must be made.

Information Provision and Acceptance

If CV surveys are to elicit useful information about willingness to pay, respondents must understand exactly what it is they are being asked to value (or vote upon) and must accept the scenario in formulating their responses. Frequently, CV surveys have provided only sketchy details about the project(s)

being valued and this calls into question the estimates derived therefrom.

Consider the following example. Suppose information is desired about individuals' willingness to pay to prevent a chemical leak into a river. Presumably, their responses would depend importantly on how long it would take for the chemical to degrade naturally in the river (if it would at all), what ecological and human health damage the chemical would do until it had degraded, and so on. Absent information about such matters, it is unreasonable to expect even very bright and well-informed respondents to place meaningful values on a program to prevent leaks.

Even if detailed information were supplied, there are limits on the ability of respondents to internalize and thus accept and proceed from the information given. It is one thing to tell respondents matter-of-factly that complete recovery will occur in, say, two years. It is another thing for them to accept this information completely and then incorporate it in their answers to difficult questions.

To return to the example above, respondents who take a pessimistic view of the probable consequences of a chemical leak are likely to report relatively high willingness to pay to prevent the contamination -- too high, in fact, if in actuality such an event had less serious effects. On the other hand, respondents with an exaggerated sense of the river's assimilative capacity or regenerative power could be expected to report a willingness to pay that understates their "true" valuation if

provided with a more complete description of likely consequences.

To repeat, even when CV surveys provide detailed and accurate information about the effects of the program being valued, respondents must accept that information in making their (hypothetical) choices. If, instead, respondents rely on a set of heuristics ("these environmental accidents are seldom as bad as we're led to believe," or "authorities almost always put too good a face on these things"), in effect they will be answering a different question from that being asked; thus, the resulting values that are elicited will not reliably measure willingness to pay.

Extent of the Market

Suits for environmental damages are brought by trustees on behalf of a legally definable group. This group limits the population that is appropriate for determining damages even though individuals outside of this group may suffer loss of passive and active use. Undersampling and even zero sampling of a subgroup of the relevant population may be appropriate if the subgroup has a predictably low valuation of the resource. For example, the authors of the CV study conducted in connection with the Nestucca oil spill limited their sample to households in Washington and British Columbia possibly because the individuals living elsewhere were presumed to have values too low to justify examination (or possibly because the sponsors of the study were agencies of the State of washington and the province of British Columbia and so defined the legally appropriate population)

(Rowe, Shaw, and Schulze, 1992).

"Warm Glow" Effects

Some critics of the CV technique (e.g., Diamond and Hausman (1992)) have observed that the distribution of responses to open-ended questions about willingness to pay often is characterized by a significant proportion of "zeros" -- people who would pay nothing for the program -- and also a number of sizable reports. This bi-modal distribution also characterizes individual giving: most of us give nothing to most charities, but give non-trivial amounts to the ones we do support (at least $10 or $20, say). This has led these critics to conclude that individuals' responses to CV questions serve the same function as charitable contributions -- not only to support the organization in question, but also to feel the "warm glow" that attends donating to worthy causes (see Andreoni (1989)). If this is so, CV responses should not be taken as reliable estimates of true willingness to pay, but rather as indicative of approval for the environmental program in question.

III. KEY ISSUES IN THE DESIGN OF CONTINGENT VALUATION INSTRUMENTS

In the course of its deliberations, the Panel discussed many issues surrounding the design of CV surveys. Here we provide our views on several issues that are especially important. In Section IV and in an Appendix to this report, we provide much greater detail on the characteristics of a valid application of

the CV method.

The Referendum Format

Considered as a survey, a CV instrument is descriptive rather than explanatory. Description may be as simple as reporting univariate averages of one kind or another, such as the percentages of those employed, seeking work, and not seeking work in the United States, the mean number of rooms occupied by American households, or the proportion of "likely" voters favoring one or another candidate in an upcoming election. A CV study seeks to find the average willingness to pay for a specific environmental improvement. Nevertheless, as will be seen later, it is often desirable to ask respondents to specify the reasons for their reported choices.

Univariate descriptive results are meaningful mainly when the alternative responses to a question are simple and can be well specified and there is a high consensus among both respondents and investigators about the precise meaning of the questions and answers. In some cases where consensus would initially not be adequate, simple definitions can be added to a questionnaire to attain satisfactory agreement -- e.g., in asking people how many rooms they have in their homes, one states whether bathrooms, basements, etc. are to be included in the count; most respondents will conform to this specification.

With questions about subjective phenomena, such as attitudes and values, treating answers as simply descriptive is seldom meaningful. Too much depends on how questions are worded, and

there is neither sufficient social consensus about precise meaning, nor an external reference to facilitate such consensus. There are many examples in the survey literature of how changes in wording or context will affect results based on questions about subjective phenomena (see Schuman and Presser (1981)). For example, in national surveys close to a quarter of the population will choose the "don't know" response to most attitude questions if it is explicitly offered; yet these same people will select a substantive alternative if "don't know" is not specifically provided, even though accepted when asserted spontaneously. More puzzlingly, a question about "forbidding" a particular action tends to elicit less agreement than a question about "not allowing" the same action, although the two questions are logically equivalent. Beyond these examples, most attitude objects are simply too complex to be summarized by a single survey question, e.g., attitudes toward abortion are too dependent on the reasons for abortion and the time in pregnancy to be adequately captured by a single question; attitudes toward "gun control" vary enormously depending on the exact framing of the issue (e.g., handguns vs. all guns, registration vs. banning, and other concrete policy distinctions).

Contingent valuation studies seek descriptive information, yet call for a response similar to those elicited by questions about subjective phenomena. Thus they risk many of the same response effects and other wording difficulties that turn up regularly in attitude surveys. Minimizing these effects presents a considerable challenge to anyone wishing to elicit reliable CV

estimates. The simplest way to approach the problem is to consider a CV survey as essentially a self-contained referendum in which respondents vote on whether to tax themselves or not for a particular purpose. Since real referenda are exposed to most of the response effects that occur with attitude surveys, and since we take the result of referenda as telling us something about "true" preferences, it is not necessary to claim they can be eliminated completely in a CV study.

The Panel is of the opinion that open-ended CV questions -- e.g., "What is the smallest sum that would compensate you for environmental damage X?" or, "What is the largest amount you would be willing to pay to avoid (or repair) environmental damage X?" -- are unlikely to provide the most reliable valuations. There are at least two reasons for this conclusion. In the first place, the scenario lacks realism since respondents are rarely asked or required in the course of their everyday lives to place a dollar value on a particular public good. Their responses to such questions are therefore likely to be unduly sensitive to trivial characteristics of the scenario presented. In the second place, an open-ended request for willingness to pay or willingness to accept compensation invites strategic overstatement. The more seriously the respondent takes the question, the more likely it is that he or she will see that reporting a large response is a costless way to make a point. Both experience and logic suggest that responses to open-ended questions will be erratic and biased.

However, the referendum format, especially when cast in the

willingness to pay mode -- "Would you be willing to contribute (or be taxed) D dollars to cover the cost of avoiding or repairing environmental damage X?" -- has many advantages. It is realistic: referenda on the provision of public goods are not uncommon in real life. There is no strategic reason for the respondent to do other than answer truthfully, although a tendency to overestimate often appears even in connection with surveys concerning routine market goods. The fact that market surveys continue to be used routinely suggests that this tendency is not a insuperable obstacle. Of course, the respondent in a CV survey understands that the referendum is hypothetical; there is no implication that the tax will actually be levied and the damage actually repaired or avoided. This suggests that considerable efforts should be made to induce respondents to take the question seriously, and that the CV instrument should contain other questions designed to detect whether the respondent has done so. Although Carson, et al. (1992), included a useful question to determine whether respondents believed the survey was biased in any direction, they did not sufficiently test whether the completeness of, and time period for, restoration stated in the survey were fully accepted by respondents. But, as far as strategic reasons go, a respondent who would not be willing to pay D dollars has no reason to answer "Yes," and a respondent who would be willing to pay D dollars has no reason to answer "No."

There are, however, several other reasons why one's response to a hypothetical referendum question might be the opposite of one's actual vote on a real ballot. On one hand, a respondent

unwilling to pay D dollars in reality might feel pressure to give the "right" or "good" answer when responding to an in-person or telephone interviewer. This could happen if the respondent believes that the interviewer would herself favor a yes answer. On the other hand, a respondent actually willing to pay the stated amount might answer in the negative for several reasons: (i) belief that the proposed scenarios distributed the burden unfairly; (ii) doubt of either the feasibility of the proposed action, so that any contribution would be wasted, or the ability of the relevant agency to carry out the action efficiently; or (iii) refusal to accept the hypothetical choice problem, because of either a generalized aversion to taxes or a view that someone else -- the "oil industry", for example -- should pay for repair or avoidance as the responsible party. The same considerations suggest that a CV instrument should include questions designed to detect the presence of these sources of bias. This is in fact often done, but we do not know how successfully.

There are two further problems that could detract from the reliability of CV responses without producing any determinate bias: (i) a feeling that one's vote will have no significant effect on the outcome of the hypothetical referendum, leading to no reply or an unconsidered one; and (ii) poor information about the damage being valued. Of course, either of these could occur in real referenda.

Here we must decide on the standard of knowledgeability of the respondents that we want to impose on a CV study. It is clear that it should be at least as high as that which the

average voter brings to a real referendum on the provision of a specific public good, but should it be higher? A "conservative" CV study, i.e., one that avoids overestimating true willingness to pay, will no doubt exceed the minimum standard of information and will also lean over backwards to avoid providing information in a way that might bias the response upwards. In particular, a conservative study will provide the respondent with some perspective concerning the overall frequency and magnitude of oil spills, the amount of money currently being spent on preventing and remedying them, the overall scale of their consequences, the peculiar features of the spill in question, and similar relevant information. Placing the choice problem in a broader context helps the respondent to arrive at a realistic or even conservative valuation.

Most of the provision of public goods in this country is decided by representatives and bureaucrats rather that by direct vote of the citizens. It is presumed that these agents are more "expert" or at least draw on more knowledge than the citizens themselves. The agents' expertise, if it really exists, is about the means and cost of providing public goods, though elected officials may sometimes be presumed to "represent" judgments of ultimate value to the citizens. Nevertheless, to increase one's confidence that a CV study is conservatively reliable, one might want to compare its outcome with that provided by a panel of experts. This will help check whether respondents and those conducting the study or studies are reasonably well-informed and well-motivated. This comparison could be made on a sample of CV

studies to give an idea of their reliability in general.

The above considerations suggest that a CV study based on the referendum scenario can produce more reliably conservative estimates of willingness to pay, and hence of compensation required in the aftermath of environmental impairment, provided that a concerted effort is made to motivate the respondents to take the study seriously, to inform them about the context and special circumstances of the spill or other accident, and to minimize any bias toward high or low answers originating from social pressure within the interview. This implies that, in the present state of the art, a reliably conservative CV study should be conducted with personal interviews of significant duration and will therefore be relatively costly. If follows therefore that, in order that the cost of the study not be disproportionately large compared to the amount of damages, the CV approach would likely be used only in relatively major spills, at least until further improvements in methodology can be developed and accepted. (A suggestion for doing so is offered in Section V.)

The referendum format offers one further advantage for CV. As we have argued, external validation of elicited lost passive-use values is usually impossible. There are however real-life referenda. Some of them, at least, are decisions to purchase specific public goods with defined payment mechanisms, e.g., an increase in property taxes. The analogy with willingness to pay for avoidance or repair of environmental damage is far from perfect but close enough that the ability of CV-like studies to predict the outcomes of real-world referenda would be useful

evidence on the validity of the CV method in general.

The test we envision is not an election poll of the usual type. Instead, using the referendum format and providing the usual information to the respondents, a study should ask whether they are willing to pay the average amount implied by the actual referendum. The outcome of the CV-like study should be compared with that of the actual referendum. The Panel thinks that studies of this kind should be pursued as a method of validating and perhaps even calibrating applications of the CV method (see Magleby, 1984).

Addressing the Embedding Problem

Perhaps the most important internal argument against the reliability of the CV approach (as against general criticisms about vagueness, lack of information, or unreality of the scenario) is the observation of the "embedding" phenomenon (see the discussion in Section II). Different but similar samples of respondents are asked about their willingness to pay for prevention of environmental damage scenarios that are identical except for their scale: different numbers of seabirds saved, different numbers of forest tracts preserved from logging, etc. It is reported that average willingness to pay is often substantial for the smallest scenario presented but is then substantially independent of the size of the damage averted, rising slightly if at all for large changes in size.

The usual interpretation proposed by critics of the CV method is that the responses are not measuring the equivalent

dollar value of the utility of the environmental assets preserved, because that would certainly be measurably larger for substantially larger programs of preservation. Instead, the fixed sum offered is the value of a feeling of having done something praiseworthy; a "warm glow" is the phrase often used.

This is potentially a very damaging criticism of the method. CV studies almost always seek to measure willingness to pay to avoid a particular incident rather than compensation that would be required for damage that has already occurred. This is because respondents are more likely to exaggerate the compensation they would require than their willingness to pay, and because the latter is expected to be less than the former and so is conservative. If reported willingness to pay accurately reflected actual willingness to pay, then, under the "warm glow" interpretation, willingness to pay might well exceed compensation required because the former contains an element of self-approbation. It might be real but not properly compensable.

Defenders of the CV approach reply to this criticism in various ways. Sometimes it is argued that the evidence used to support "embedding" simply indicates diminishing marginal utility of the asset in question. In many cases, however, the constancy or near-constancy of willingness to pay does not appear consistent with the large reported amounts for the first small increment of environmental preservation.

A second defense of CV against the embedding phenomenon is that CV questions have to be posed carefully and in context. It is argued that carelessly formulated CV instruments leave

respondents with the impression that they are being asked, "Would you pay $X to avert a certain small environmental harm?" In a very large population of birds, the death of 1,000 is not seen as noticeably different from the death of 100,000 -- and may not actually be very different -- so that respondents simply answer the question just asked.

This second response leads to the obvious question: how should a CV instrument be framed to elicit an answer that responds to the precise scenario and not to a generalized "warm glow" effect? We must reject one possible approach, that of asking each respondent to express willingness to pay to avert incidents of varying sizes; the danger is that embedding will be forcibly avoided, still without realism. This issue is best considered as part of the broader question: How much context about the incident itself and about the respondent's circumstances and choices should be included in the CV instrument?

We are recommending a high standard of richness in context to achieve a realistic background. Our proposed guidelines regarding this issue are embodied in Section IV below.

Time Dimension of Passive Use Losses

Typically, environmental damages from oil spills or similar accidents are severe for some period of time -- weeks, months, or sometimes a few years -- and gradually are reduced by natural forces and human efforts to a low or possibly even zero steady state level. In some circumstances, passive-use losses derive

only or mostly from the steady state conditions; thus, if passive use value derives from species diversity, even a considerable loss of birds or mammals which does not endanger any species will give rise to no loss. If, on the contrary, considerable passive-use value is attached to the interim state of the natural resource, then respondents have to do a very difficult present value calculation properly to compute their current willingness to pay for the difference between the fully restored state of the resource and the actual state as the level of restoration varies over time. CV surveys accordingly have to be carefully designed to allow respondents to differentiate interim from steady state passive-use loss, and, if there is interim passive-use loss, to report its present value correctly.

It is reasonable to assume that interim passive-use values are additive over time. Hence, we need a calculation of present values of the interim losses. The discounting and the estimation of the rate of recovery of the resource should be done by technical experts and not by the respondents, who are unlikely to handle these tasks adequately. Respondents should be asked only their willingness to pay to eliminate the difference between some partially restored level of the resource and the pristine state for a specific period of time, say a year, on the assumption that after that time full restoration is assured. Technical experts would estimate how the state of the resource will vary from year to year as the restoration takes place. The technical information about the state of the resource, together with the respondent's assessments of the flow valuation of the resource,

can be used to construct a time series of passive-use losses which can be discounted to the present at an appropriate rate of interest to determine the present value of the damages.

IV. SURVEY GUIDELINES

In this section we try to lay down a fairly complete set of guidelines compliance with which would define an ideal CV survey. A CV survey does not have to meet each of these guidelines fully in order to qualify as a source of reliable information to a damage assessment process. Many departures from the guidelines or even a single serious deviation would, however, suggest unreliability *prima facie*. To preserve continuity, we give only a bald list of guidelines here. They are repeated together with further explanatory comments in the Appendix to this Report.

GENERAL GUIDELINES

- Sample Type and Size: Probability sampling is essential for a survey used for damage assessment.[1] The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

---

[1] This need not preclude use of less adequate samples, including quota or even convenience samples, for preliminary testing of specific experimental variations, so long as order of magnitude differences rather than univariate results are the focus. Even then, obvious sources of bias should be avoided (e.g., college students are probably too different in age and education from the heterogeneous adult population to provide a trustworthy basis for wider generalization).

- Minimize Nonresponses: High nonresponse rates would make the survey results unreliable.

- Personal Interview: The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision.

- Pretesting for Interviewer Effects: An important respect in which CV surveys differ from actual referenda is the presence of an interviewer (except in the case of mail surveys). It is possible that interviewers contribute to "social desirability" bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CV studies should incorporate experiments that assess interviewer effects.

- Reporting: Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties (see

Carson et al. (1992), for an example of good practice in inclusion of questionnaire and related details; as of this date, however, the report has not been available publicly and the data have not been archived for open use by other scholars).

- Careful Pretesting of a CV Questionnaire: Respondents in a CV survey are ordinarily presented with a good deal of new and often technical information, well beyond what is typical in most surveys. This requires very careful pilot work and pretesting, plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

GUIDELINES FOR VALUE ELICITATION SURVEYS

The following guidelines are met by the best CV surveys and need to be present in order to assure reliability and usefulness of the information that is obtained.

- Conservative Design: Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred. A conservative design increases the reliability of the estimate by eliminating extreme responses that can enlarge estimated values wildly and implausibly.

- Elicitation Format: The willingness to pay format should be

used instead of the compensation required because the former is the conservative choice.

- Referendum Format: The valuation question should be posed as a vote on a referendum.

- Accurate Description of the Program or Policy: Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

- Pretesting of Photographs: The effects of photographs on subjects must be carefully explored.

- Reminder of Undamaged Substitute Commodities: Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resource. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

- Adequate Time Lapse from the Accident: The survey must be conducted at a time sufficiently distant from the date of the environmental insult that respondents regard the scenario of complete restoration as plausible. Questions should be included to determine the state of subjects'

beliefs regarding restoration probabilities.

- Temporal Averaging: Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial time trend in the responses would cast doubt on the "reliability" of the finding.

- "No-answer" Option: A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation (referendum) question. Respondents who choose the "no-answer" option should be asked nondirectively to explain their choice. Answers should be carefully coded to show the types of responses, for example: (i) rough indifference between a yes and a no vote; (ii) inability to make a decision without more time or more information; (iii) preference for some other mechanism for making this decision; and (iv) bored by this survey and anxious to end it as quickly as possible.

- Yes/no Follow-ups: Yes and no responses should be followed up by the open-ended question: "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses, for example: (i) It is (or isn't) worth it; (ii) Don't know; or (iii) The oil companies should pay.

- Cross-tabulations: The survey should include a variety of

other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of willingness to pay broken down by these categories. Among the items that would be helpful in interpreting the responses are:

- Income
- Prior Knowledge of the Site
- Prior Interest in the Site (Visitation Rates)
- Attitudes Toward the Environment
- Attitudes Toward Big Business
- Distance to the Site
- Understanding of the Task
- Belief in the Scenarios
- Ability/Willingness to Perform the Task

° Checks on Understanding and Acceptance: The above guidelines must be satisfied without making the instrument so complex that it poses tasks that are beyond the ability or interest level of many participants.

GOALS FOR VALUE ELICITATION SURVEYS

The following items are not adequately addressed by even the best CV surveys. In the opinion of the Panel, these issues will need to be convincingly dealt with in order to assure the reliability of the estimates.

° Alternative Expenditure Possibilities: Respondents must be

reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario.

- Deflection of Transaction Value: The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated. It is possible that the referendum format limits the "warm glow" effect, but until this is clear the survey design should explicitly address this problem.

- Steady State or Interim Losses: It should be made apparent that respondents can distinguish interim from steady-state losses.

- Present Value Calculations of Interim Losses: It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

- Advance Approval: Since the design of the CV survey can have a substantial effect on the responses, it is desirable

that -- if possible -- critical features be preapproved by both sides in a legal action, with arbitration and/or experiments used when disagreements cannot be resolved by the parties themselves.

- Burden of Proof: Until such time as there is a set of reliable reference surveys, the burden of proof of reliability must rest on the survey designers. They must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. Specifically, if a CV survey suffered from any of the following maladies, we would judge its findings "unreliable":

  -- A high nonresponse rate to the entire survey instrument or to the valuation question.

  -- Inadequate responsiveness to the scope of the environmental insult.

  -- Lack of understanding of the task by the respondents.

  -- Lack of belief in the full restoration scenario.

  -- "Yes" or "no" votes on the hypothetical referendum that are not followed up or explained by making

reference to the cost and/or the value of the program.

- Reliable Reference Surveys: In order to alleviate this heavy burden of proof, we strongly urge the government to undertake the task of creating a set of reliable reference surveys that can be used to interpret the guidelines and also to calibrate surveys that do not fully meet the conditions.

V. RECOMMENDATIONS FOR FUTURE RESEARCH

The Panel's major research recommendation goes toward a drastic reform of the CV procedure, extending beyond the guidelines suggestion in Section IV.

The problem of estimating the demand for highly innovative commercial products, including some that have not yet actually been produced, is much like the problem faced in CV research. It is the problem of estimating willingness to pay for a necessarily unfamiliar product. The field of market research has developed methods -- "conjoint analysis," for example -- that are very similar to the CV approach. (One important difference is that a new product may eventually reach the market, and projections of expected sales can be checked. Survey responses are usually found to be moderate overestimates of actual willingness to pay.) Practitioners have found that survey methods are better at estimating relative demand than absolute demand. There is an anchoring problem, even with private goods -- that is, absolute

willingness to pay is hard to pin down. This leads to the following suggestion.

The federal government should produce standard damage assessments for a few specific reference oil spills, either hypothetical or actual, ranging from small to large. These standard valuations could be generated by any method. One possibility would be through a jury of experts. Such a jury of experts might wish to conduct a series of CV studies, satisfying the guidelines laid out above. These CV studies would be inputs into the jury process, to be combined with other information and expert judgment. Once these benchmarks were available, they could serve as reference points for later CV studies. When a damage assessment is required, surveys could be used to elicit answers to questions like: "Would you pay (much more, more, about the same, less, much less) to prevent this spill than you would to prevent Standard Spill A?" "Would you pay an amount to avoid this spill that is between the amounts you would pay to avoid Standard Spill B and Standard Spill C? If so, is the amount much closer to B than C, closer to B than C, halfway between B and C, closer to C than B, much closer to C than B?" These questions presumably would not be asked so schematically. Responses to such a study could then serve as one reliable source of information in the damage assessment.

We recognize that this technique would require that respondents be made familiar with the reference spills as well as the particular spill whose damage is being assessed. We expect that the additional effort would be more than offset by the

greater simplicity and reliability in estimating relative willingness to pay.

This possibility suggests a slightly more radical extension of the CV method. Respondents could be asked to compare their willingness to pay to avoid a specific case of environmental damage to their willingness to pay for a range of fairly familiar private goods. It would no doubt be best if the private goods were to bear some similarity to the environmental good in question, but that is not necessary. The anchoring purpose would be served if respondents could measure their willingness to pay in units of articles of clothing or small household appliances forgone.

This latter is a suggestion for research in the CV method, not necessarily a recommendation for current practical use.

The guidelines proposed in Section IV themselves suggest areas for further research, this time within the contingent valuation community. In particular, we emphasize the urgency of studying the sensitivity of willingness to pay responses to the number and extent of budgetary substitutes mentioned in survey instruments (that is, reminders of other things on which respondents could spend their money). In such research it would be helpful if parallel studies were conducted on the sensitivity of stated intentions to buy ordinary market goods -- both familiar and unfamiliar -- to reminders of alternative uses of those resources. The point is to discover the extent to which the valuation of environmental public goods is intrinsically more difficult than similar exercises with respect to market goods.

A closely-related line of research is the sensitivity of responses in CV surveys to the number and extent of undamaged substitute commodities mentioned explicitly in the survey instrument (miles of nearby shoreline, miles of shoreline elsewhere, similarity for animal or bird life, alternative recreation possibilities and so on). This could be extended to variations in the way in which the budget constraint is presented to respondents. Here again, comparisons with market goods would be useful.

Finally, having urged that the availability of a no-vote option is an important component of the ability of the CV technique to mimic an actual referendum, we recommend further research into alternative ways of presenting and interpreting the no-vote option. In this respect, too, comparative studies with familiar public and private goods (local parks, school facilities, housing for the homeless, food distributions) would be enlightening. Real referenda always allow the option of not voting, in a natural way. CV studies have to achieve the same result more deliberately, so there is a need to know if the precise formulation matters very much to the result.

VI. CONCLUSIONS AND RECOMMENDATIONS

The Panel starts from the premise that passive-use loss -- interim or permanent -- is a meaningful component of the total damage resulting from environmental accidents. A problem arises because passive-use losses have few or no overt behavioral consequences. The faintness of the behavioral trail means that a

well-designed and adequately sensitive measuring instrument is needed to substitute for conventional observations of behavior. In particular, can the CV method provide a sufficiently reliable estimate of total loss -- including passive-use loss -- to play a useful role in damage assessment?

It has been argued in the literature and in comments addressed to the Panel that the results of CV studies are variable, sensitive to details of the survey instrument used, and vulnerable to upward bias. These arguments are plausible. However, some antagonists of the CV approach go so far as to suggest that there can be no useful information content to CV results. The Panel is unpersuaded by these extreme arguments.

In Section IV above, we identify a number of stringent guidelines for the conduct of CV studies. These require that respondents be carefully informed about the particular environmental damage to be valued, and about the full extent of substitutes and undamaged alternatives available. In willingness to pay scenarios, the payment vehicle must be presented fully and clearly, with the relevant budget constraint emphasized. The payment scenario should be convincingly described, preferably in a referendum context, because most respondents will have had experience with referendum ballots with less-than-perfect background information. Where choices in formulating the CV instrument can be made, we urge they lean in the conservative direction, as a partial or total offset to the likely tendency to exaggerate willingness to pay.

The Panel concludes that under those conditions (and others

specified above), CV studies convey useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis for new and innovative products and the assessment of other damages normally allowed in court proceedings. As in all such cases, the more closely the guidelines are followed, the more reliable the result will be. It is not necessary, however, that every single injunction be completely obeyed; inferences accepted in other contexts are not perfect either.

Thus, the Panel concludes that CV studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values. To be acceptable for this purpose, such studies should follow the guidelines described in Section IV above. The phrase "be the starting point" is meant to emphasize that the Panel does not suggest that CV estimates can be taken as automatically defining the range of compensable damages within narrow limits. Rather, we have in mind the following considerations.

The Panel is persuaded that hypothetical markets tend to overstate willingness to pay for private as well as public goods. The same bias must be expected to occur in CV studies. To the extent that the design of CV instruments makes conservative choices when alternatives are available, as urged in Section IV, this intrinsic bias may be offset or even over-corrected. All surveys of attitudes or intentions are bound to exhibit sensitivity of response to the framing of questions and the order

in which they are asked. No automatic or mechanical calibration of responses seems to be possible.

The judicial process must in each case come to a conclusion about the degree to which respondents have been induced to consider alternative uses of funds and take the proposed payment vehicle seriously. Defendants will argue that closer attention to substitute commodities would have yielded lower valuations. Trustees will argue that they have already leaned over backwards to ensure conservative responses. Judges and juries must decide as they do in other damage cases. The Panel's conclusion is that a well-conducted CV study provides an adequately reliable benchmark to begin such arguments. It contains information that judges and juries will wish to use, in combination with other evidence, including the testimony of expert witnesses.

The Panel's second conclusion is that the appropriate federal agencies should begin to accumulate standard damage assessments for a range of oil spills, as described in Section V. That process should further improve the reliability of CV studies in damage assessment. It should thus contribute to increasing the accuracy and reducing the cost of subsequent damage assessment cases. In that sense, it can be regarded as an investment.

The proposals for further research outlined in Section V are an integral part of our recommendations. The Panel believes that the suggestions put forward there could lead to more reliable and less controversial damage assessment at reduced cost. It is not to be expected that controversy will disappear, however. There

will always be controversy where intangible losses have to be evaluated in monetary terms.

APPENDIX

GENERAL GUIDELINES

- Sample Type and Size: Probability sampling is essential for a survey used for damage assessment.[1] The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

If a single dichotomous question of the yes-no type is used to elicit valuation responses, then a total sample size of 1000 respondents will limit sampling error to about 3% plus or minus on a single dichotomous question, assuming simple random sampling. However, this or any other sample size needs to be reconceptualized for three reasons. First, if face-to-face interviewing is used, as we suggest above, clustering and stratification must be taken into account. Second, if dichotomous valuation questions are used (e.g., hypothetical referenda), separate valuation amounts must be asked of random sub-samples and these responses must be unscrambled econometrically to estimate the underlying population mean or median. Third, in order to incorporate experiments on interviewer and wording effects, additional random sub-sampling

---

[1] This need not preclude use of less adequate samples, including quota or even convenience samples, for preliminary testing of specific experimental variations, so long as order of magnitude differences rather than univariate results are the focus. Even then, obvious sources of bias should be avoided (e.g., college students are probably too different in age and education from the heterogeneous adult population to provide a trustworthy basis for wider generalization).

is required. For all these reasons, it will be important to consult sampling statisticians in the design of a CV survey intended for legal or policy-making purposes.

- Minimize Nonresponses: High nonresponse rates would make the survey results unreliable.

To the extent that a CV study is expected to represent the adult population of the United States or a portion of it, minimizing both sample non-response and item non-response are important. The former is unlikely to be below 20% even in very high quality surveys; the latter has also been large in some CV surveys because of the difficulty of the task respondents are being asked to perform. These sources of potential bias can be partially justified on the grounds that they also occur with official referenda, in both cases with the loss especially of the least educated parts of the population. The further reduction of the final sample by elimination of "protest zeros," "unrealistic high values," and other problematic responses may lead to effective final total response rates so low as to imply that the survey population consists of interested and specially instructed quasi-experts. This consideration reinforces the desirability of combining a reasonable response rate with a high but not forbidding standard of information, as discussed in Section III above.

- Personal Interview: The Panel believes it unlikely that

reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision.

Assuming a CV survey is to represent a natural population, such as all adults in the United States, or those in a single urban area or a state, it is desirable that it be carried out using either face-to-face or telephone interviews. Mail surveys typically employ lists that cover too small a part of the population (e.g., samples based on telephone directories omit approximately half the U.S. population because of non-listed numbers, incorrect numbers, and non-phone households), and then miss another quarter or more of the remainder through non-response. In addition, since the content of a mail questionnaire can be reviewed by targeted respondents before deciding to return it, those most interested in a natural resource issue or in one side or the other can make their decision on that basis. It is also impossible using mail surveys to guarantee random selection within households or to confine answering to a single respondent, and it is difficult (though not impossible) to control question-order effects. Thus, mail surveys should be used only if another supplementary method can be employed to cross-validate the results on a random sub-sample of respondents.

The choice between telephone and face-to-face administration is less clear. Face-to-face surveys offer practical advantages in maintaining respondent motivation and allowing use of graphic

supplements. Both coverage and response rates are also usually somewhat higher than with telephone surveys. However, telephone surveys can cut interviewing costs by between a third and a half; for CV purposes, it may be a disadvantage that most survey investigators believe telephone interviews need to be kept shorter in length than face-to-face interviews because respondent attention and cooperation are more difficult to maintain. In addition, random-digit-dial telephone surveys approximate simple random sampling. Face-to-face surveys must be based on cluster sampling and, therefore, the results provide less precise estimates than do telephone surveys of the same size.

- Pretesting for Interviewer Effects: An important respect in which CV surveys differ from actual referenda is the presence of an interviewer (except in the case of mail surveys). It is possible that interviewers contribute to "social desirability" bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CV studies should incorporate experiments that assess interviewer effects.

To test for interviewer effects, two modifications might be made to a standard face-to-face CV survey. In one variant on current practice, respondents would stop when they come to the valuation question, write their "vote" on a ballot, and fold and deposit it in a sealed box. However, since this practice would not mimic the complete anonymity of the voting booth, for a sub-

sample of respondents a second modification should be made. Respondents would be allowed to mail their "ballots" in unmarked envelopes directly to the survey organization, even though that will preclude any but the simplest analysis of responses. Tests of the effect of both these modifications of current practice will indicate whether they are needed routinely or whether at least some calibration should be introduced to compensate for interviewer effects. (The more modest of these proposed modifications -- a simulated ballot box, or even voting on a portable computer -- has few if any disadvantages and might be made standard if it shows any reliable departure at all from answers given orally to the interviewer.)

- Reporting: Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties (see Carson et al. (1992), for an example of good practice in inclusion of questionnaire and related details; as of this date, however, the report has not been available publicly and the data have not been archived for open use by other scholars).

- Careful Pretesting of a CV Questionnaire: Respondents in a CV survey are ordinarily presented with a good deal of new and often technical information, well beyond what is typical in most surveys. This requires very careful pilot work and pretesting, plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

Parenthetically, the claim sometimes made by CV proponents that particular methods of piloting, such as focus groups, are essential should be viewed with skepticism, since these claims are unsupported by any systematic evidence. Nor is it clear that what are called "state-of-the-art" CV surveys constitute something entirely new or different from other types of serious survey investigations. Thus, although evidence that questionnaire development has been carried out carefully is certainly important, it cannot be taken as a self-sufficient basis of validity -- the more so because we know that many people will answer survey questions without apparent difficulty, even when they do not understand them well. A way of reducing pressure to give answers of questionable meaningfulness would be to provide respondents an explicit "no opinion" type of alternative when a key valuation question is posed.

GUIDELINES FOR VALUE ELICITATION SURVEYS

The following guidelines are met by the best CV surveys and need to be present in order to assure reliability and usefulness

of the information that is obtained.

- Conservative Design: Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred. A conservative design increases the reliability of the estimate by eliminating extreme responses that can enlarge estimated values wildly and implausibly.

- Elicitation Format: The willingness to pay format should be used instead of compensation required because the former is the conservative choice.

In experimental settings, the gap between stated intentions to support a particular referendum and actual behavior in the voting booth can be very great (see Magleby, 1984). This gap might be treated by "calibration" if there were historical data on the relationship between such intentions and behavior. Unfortunately, we are aware of no data that is close enough to the CV context that could be used to calibrate CV responses. In the absence of historical data that can be used to calibrate the intentions reported in the CV surveys, the survey instrument has to be designed with extraordinary care so that it can stand on its own.

- Referendum Format: The valuation question should be posed as a vote on a referendum.

As is now generally recognized by most CV proponents, asking respondents to give a dollar valuation in response to an open-ended question presents them with an extremely difficult task. At the same time, CV proponents also recognize that presenting respondents a set of dollar amounts from which they are to choose is likely to create anchoring and other forms of bias. Thus, we recommend as the most desirable form of CV elicitation the use of a dichotomous question that asks respondents to vote for or against a particular level of taxation, as occurs with most real referenda. As already noted, such a question form also has advantage in terms of incentive compatibility. (If a double-bounded dichotomous choice or some other question form is used in order to obtain more information per respondent, experiments should be developed to investigate biases that may be introduced.)

- Accurate Description of the Program or Policy: Adequate information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

Ideally a CV survey would elicit attitudes toward three alternative (future) recovery scenarios: (A) "immediate" restoration, (b) accelerated restoration, and (c) natural restoration._ Damages would be the difference between (a) and (b) on the assumption that accelerated restoration is provided by the responsible party. Unfortunately, respondents may not find

"immediate" restoration very plausible and they may resist the notion that they should be expected to contribute to accelerated restoration when it is an oil company that is at fault. If respondents are unable or unwilling to deal hypothetically with the most relevant "clean-up" scenarios, alternative "prevention" scenarios will have to be used in the survey instrument. For example, respondents may be asked to vote for a referendum that offers reduced risk of another spill for a specified period of time.[2] The weaker is the linkage between the "prevention" scenarios and the "clean-up" scenarios, the more unreliable are the survey results. Rhetorically: Is a decade of prevention equal in value to the difference in value between accelerated and immediate clean-up?

- Pretesting of Photographs: The effects of photographs on subjects must be carefully explored.

One effective means for conveying information and holding interest in a CV interview has been the use of large and impressive photographs. However, this technique is a two-edged sword because the dramatic nature of a photograph may have much more emotional impact than the rest of the questionnaire. Thus it is important that photographs be subjected to even more careful assessment than verbal material if the goal is to avoid

---

[2] As in the survey actually performed by the State of Alaska after the *Valdez* spill (See Carson et al. (1992)).

bias in presentation.[3]

- Reminder of Undamaged Substitute Commodities: Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resource. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

- Adequate Time Lapse from the Accident: The survey must be conducted at a time sufficiently distant from the date of the environmental insult that respondents regard the scenario of complete restoration as plausible. Questions should be included to determine the state of subjects' beliefs regarding restoration probabilities.

Survey respondents who would not suffer interim passive-use loss may not regard full restoration as very plausible; therefore, they may report substantial passive-use loss even if told that full restoration in some reasonable amount if time is certain. Misunderstanding of the restoration probability is most acute when the accident has recently occurred and before any substantial restoration takes place. It would be ideal to assess steady state passive-use loss after natural and human restoration

[3] Failure to test the effects of photographs on responses is one shortcoming of Carson et al. (1992).

is complete or nearly so, since then presumably respondents would believe in the restoration. If that is not a possibility, surveys might be conducted over time until the reported willingness to pay settles down (assuming that it does), as the respondents come to believe more and more in the probable success of the restoration effort. Alternatively, respondents might be asked to value a menu of alternative possible scenarios, without being told explicitly which is applicable for the environmental insult under study. The menu should be designed to force them to consider the difference between interim and steady-state passive-use value.

- Temporal Averaging: Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial time trend in the responses would cast doubt on the "reliability" of the finding.

- "No-answer" Option: A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation (referendum) question. Respondents who choose the "no-answer" option should be asked nondirectively to explain their choice. Answers should be carefully coded to show the types of responses, for example: (i) rough indifference between a yes and a no vote; (ii) inability to make a decision without more time or more information; (iii) preference for some other mechanism

for making this decision; and (iv) bored by this survey and anxious to end it as quickly as possible.

- Yes/no Follow-ups: Yes and no responses should be followed up by the open-ended question: "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses, for example: (i) It is (or isn't) worth it; (ii) Don't know; or (iii) The oil companies should pay.

- Cross-tabulations: The survey should include a variety of other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of willingness to pay broken down by these categories. Among the items that would be helpful in interpreting the responses are:

  Income
  Prior Knowledge of the Site
  Prior Interest in the Site (Visitation Rates)
  Attitudes Toward the Environment
  Attitudes Toward Big Business
  Distance to the Site
  Understanding of the Task
  Belief in the Scenarios
  Ability/Willingness to Perform the Task

We believe that these cross tabulations will prove useful in interpreting and lending credibility to the responses and

possibly also in forming adjustments that can enhance reliability.

- Checks on Understanding and Acceptance: The above guidelines must be satisfied without making the instrument so complex that it poses tasks that are beyond the ability or interest level of many participants.

Since CV interviews often present information that is new to respondents, the questionnaire should attempt at the end to determine the degree to which respondents accept as true the descriptions given and assertions made prior to the valuation question. Such an inquiry should be carried out in detail but non-directively, so that respondents feel free to reject any part of the information they were given at earlier points.

GOALS FOR VALUE ELICITATION SURVEYS

The following items are not adequately addressed by even the best CV surveys. In the opinion of the Panel, these issues will need to be convincingly dealt with in order to assure the reliability of the estimates.

- Alternative Expenditure Possibilities: Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should

be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario.

Consumers can be expected to make expenditure decisions that are adequately sensitive to other expenditure possibilities with which they are familiar. But environmental referenda of the type presented in CV surveys are unfamiliar and respondents may not be aware of the large set of other expenditure possibilities that might be offered in future CV surveys or future referenda. Unless informed otherwise, respondents may suppose that there is only one environmental scenario that will ever be offered and they may overspend on it.

It is not at all clear how exhaustive should be the list of alternative public goods that are explicitly presented. If the list is too brief, overspending can be expected. If the list is too long, respondents will be encouraged to spread expenditures to public goods for which there is not adequate total demand and which therefore cannot really be offered to them. Also, if the list gets large enough to encompass a significant fraction of income, the gap between willingness to pay and willingness to accept may widen.

It is also not clear what form the reminder should take. It does not seem enough merely to list other environmental goods since respondents would then have to guess the level of expenditure that would be necessary to pay for the alternatives.

The survey should probably include some statement about the price of the alternatives, for example, the per capita expenditure that would be required to provide the items.

- Deflection of Transaction Value: The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated. It is possible that the referendum format limits the "warm glow" effect, but until this is clear the survey design should explicitly address this problem.

Economic models of consumer behavior generally are based on the assumption that value derives from the goods and services that are consumed, not from the process by which these goods are allocated. But happiness that derives from charitable giving may come mostly from the act of giving rather from the material changes that follow from the gift. To give another example, consumers may get pleasure from the act of shopping as well as from ownership of the goods they purchase. Words that might be useful to distinguish between these utility-producing events are "consumption value" and "transaction value," the latter referring to the process or transaction that establishes ownership.

We do not question the validity of "transaction value" or differentiate it from "consumption value" as far as damage assessment is concerned. But for both forms of value,

respondents need to be thinking clearly about the substitutes, since the closer are the substitutes the less the damage that is done. In the case of "transaction value," there are many close substitutes to cleaning up oil spills since there are many other charitable activities that can generate the same "warm glow" and there are many other ways to express hostility toward big business and modern technology.

- Steady State or Interim Losses: It should be made apparent that respondents can distinguish interim from steady-state losses.

The quality of any natural resource varies daily and seasonally around some "equilibrium" or "steady state" level. Active-use value of a resource depends on its actual state at the time of use (and at other times), not on its equilibrium. But passive-use value of a natural resource may derive only or mostly from its steady state and not from its day-to-day state. If so, full restoration at some future date eliminates or greatly reduces passive-use loss. Surveys accordingly need to be carefully designed to allow respondents to differentiate interim from steady state passive-use loss.

- Present Value Calculations of Interim Losses: It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

As discussed in Section III above, the time profile of restoration following an accident potentially is an important determinant of active-use loss and interim passive-use loss, but respondents may have little ability to distinguish between and to evaluate different profiles.

- Advance Approval: Since the design of the CV survey can have a substantial effect on the responses, it is desirable that -- if possible -- critical features be preapproved by both sides in a legal action, with arbitration and/or experiments used when disagreements cannot be resolved by the parties themselves.

- Burden of Proof: Until such time as there is a set of reliable reference surveys, the burden of proof of reliability must rest on the survey designers. They must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. Specifically, if a CV survey suffered from any of the following maladies, we would judge its findings "unreliable":

  -- A high nonresponse rate to the entire survey instrument or to the valuation question.

  -- Inadequate responsiveness to the scope of the environmental insult.

-- Lack of understanding of the task by the respondents.

-- Lack of belief in the full restoration scenario.

-- "Yes" or "no" votes on the hypothetical referendum that are not followed up or explained by making reference to the cost and/or the value of the program.

- Reliable Reference Surveys: In order to alleviate this heavy burden of proof, we strongly urge the government to undertake the task of creating a set of reliable reference surveys that can be used to interpret the guidelines and also to calibrate surveys that do not fully meet the conditions.

Table of References

Andreoni, James; "Giving With Impure Altruism: Applications to Charity and Ricardian Equivalence;" Journal of Political Economy 97 (1989); pp. 1447-1458.

Bishop, Richard C., and Thomas A. Heberlien; "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics 61 (1979); 926-930.

Carson, Richard T, Robert Cameron Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud; "A Contingent Valuation Study of Lost Passive Use Values Resulting from the *Exxon Valdez* Oil Spill;" A Report for the Attorney General of the State of Alaska; November 19, 1992.

Cummings, Ronald G., and Glenn W. Harrison; "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible;" Department of Economics, University of New Mexico, submitted to Office of General Counsel, National Oceanic and Atmospheric Administration, 1992, 18 pp.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and K. Nicole Wilson; "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability;" Paper presented at the Cambridge Economics, Inc., Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Diamond, P.A., and J.A. Hausman; "On Contingent Valuation Measurement of Nonuse Values;" Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Diamond, P.A., J.A. Hausman, G.K. Leonard, and M.A. Denning; "Does Contingent Valuation Measure Preferences? Experimental Evidence;" Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Dickie, Mark, Ann Fisher, and Shelby Gerking; Market Transactions and Hypothetical Demand Data: A Comparative Study; Journal of American Statistical Association, Vol. 82, March 1987, pp. 69-75.

Duffield, John W., and David A. Patterson; "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers;" Paper presented at the annual meeting of the American Economic Association; New Orleans, January 1991.

Kahneman, Daniel; "Comments" in Valuing Environmental Goods, edited by Ronald G. Cummings, David S. Brookshire and William D. Schulze; Totowa, N.J.: Rowman and Allanheld, 1986.

Kahneman, Daniel, and Jack Knetch; "Valuing Public Goods: The Purchase of Moral Satisfaction;" 22 JEEM 57-70; 1992.

Kemp, M.A., and C. Maxwell; "Exploring a Budget Context for Contingent Valuation Estimates;" Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Magleby, David B.; Direct Legislation, Johns Hopkins Press, 1984.

Mitchell, Robert Cameron, and Richard T. Carson; *Using Surveys to Value Public Goods: The Contingent Valuation Method*; Resources for the Future; Washington, DC, 1989; 499 pp.

Rowe, Robert D., W. Douglass Shaw, William Schulze; "*Nestucca* Oil Spill;" Chapter 20 in Natural Resource Damages: Law and Economics (eds. Kevin M. Ward and John W. Duffield); New York: John Wiley & Sons; 1992.

Schuman, Howard, and Stanley Presser; Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context; New York: Academic Press, 1981.

Seip, Kalle, and Jon Strand; Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment; Paper prepared for the SAF Center for Applied Research, Department of Economics, University of Oslo, 26 pp.

State of Ohio v. Department of the Interior, 880 F.2d 432 (D.C. Cir. 1989)

References

Balson, William E., Richard T. Carson, Michael B. Conaway, Baruch Fischhoff, W. Michael Hanemann, Annette Hulse, Raymond J. Kopp, Kerry Martin, Robert C. Mitchell, John Molenar, Stanley Presser, and Paul A. Ruud. 1990. Development and Design of a Contingent Valuation Survey for Measuring the Public's Value for Visibility Improvements at the Grand Canyon National Park. Draft report (September).

Bennett, J. W. 1984. "Using Direct Questioning to Value the Existence Benefits of Preserved Natural Areas." Australian Journal of Agricultural Economics 28 (August/December): 136-152.

Bockstael, Nancy E. and Kenneth E. McConnell. 1991. "The Demand for Quality Differentiated Goods: A Synthesis." Working Paper, Department of Agricultural and Resource Economics, University of Maryland, College Park, Md. (January).

Chestnut, L. G. and R. D. Rowe. 1990. Preservation Values for Visibility Protection at the National Parks. Report prepared for U. S. Environmental Protection Agency and National Park Service (Boulder, Co.: RCG/Hagler, Bailly, Inc).

Hanemann, W. Michael. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" American Economic Review (in press).

Knetsch, Jack L. 1990. "Environmental Policy Implications of Disparities Between Willingness to Pay and Compensation Demanded Measures of Values." Journal of Environmental Economics and Management 18: 227-237.

Larson, Douglas M. 1990. "Measuring Willingness to Pay for Nonmarket Goods." Paper presented at American Agricultural Economics Association Meetings, Vancouver, British Columbia (August 4-8).

Mitchell, R. C. and R. T. Carson. 1989. Using Surveys to Value Public Goods--The Contingent Valuation Method (Washington, D.C.: Resources for the Future, Inc.)