DEPARTMENT OF ENERGY/ENVIRONMENTAL PROTECTION AGENCY

WORKSHOP ON USING CONTINGENT VALUATION

TO MEASURE NON-MARKET VALUES

* * *

THURSDAY, MAY 19, 1994

* * *

HYATT DULLES

2300 DULLES CORNER BOULEVARD

CONCORD A AND B

HERNDON, VIRGINIA

* * *

A meeting of the DOE/EPA Workshop on Using Contingent Valuation to Measure Non-Market Values was held on Thursday, May 19, 1994, commencing at 8:47 a.m., Dr. Paul Portney, presiding.

DISCLAIMER

Although the information in this report has been funded in part by the United States Environmental Protection Agency, it does not necessarily reflect the views of the Agency and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

C O N T E N T S

Session 1

Session 2

Session 3

Dinner Speaker

P R O C E E D I N G S

(8:47 a.m.)

DR. BJORNSTAD: My name is David Bjornstad. As Chair of this event, I would like to welcome you on behalf of the Oak Ridge National Laboratory.

I would also like to spend just a moment in making a couple of announcements and recognizing a couple of people before we get going with this. As you have noticed, this is a complicated event put on.

And, I would like to thank Jim Cohn, who served as my Co-Chair; Cynthia Moody, who is the administrative person for the event; Linda Hampton, who helped Cynthia, and Joy Lee of our business office.

And, I would like to ask you if you have any problems at all to please see them at the outside desk. I have some telephone numbers in which you can take telephone calls or fax messages.

I'm sorry that Leslie Ackers, who spoke with several of you regarding financial matters, wasn't able to make it

today. And, I would like everyone to know that the telephones and restrooms are the second right. Take two rights as you go out the door.

I have also received a great deal of help as we have put this together from a number of individuals and particularly from Paul Portney and Peter Diamond, Terry Smith and Bob Shelton. I am very grateful for the help I've gotten from them.

And, I'm also grateful for the support we got from the Department of Energy and from the EPA. If their representatives have a few words to say to you as we begin today -- they gave us a great deal of support and a great deal of freedom with the agenda to pursue for this conference.

In particular, Don Rosenthal and Howard Bruenspeck with DOE helped us out; and, Mary Jo Kiley. They have been supportive. They also provided the resources that have made this possible. And, for that we are very happy.

As you know, the theme that we have settled on for this

is a research agenda or research questions. And, for that we want to encourage as broad as possible participation.

We have prepared a set of draft papers. I hope everyone has received a set. Some of those have been updated. And, make sure that you stop by the desk and have the latest papers that are available.

I want to encourage you to provide written comments or other comments from the audience so that we may have the benefit of your thoughts. And, our goal will be to put out a published conference volume. So, please give people your best ideas along the way.

We are going to have a transcript of the meeting. And, as a consequence, anyone that speaks will need to speak into a microphone so that they can be recorded. We have two microphones here and we have microphones in the front.

I will make sure that this document is placed in the reading rooms at EPA and DOE. And, we will make electronic copies.

It's going to be a big document. And, I'm not going to

attempt to send it to everyone. But, if you have a need for it, we will make sure you can get a disk of it or something like this.

Finally, we are going to prepare some summary statements of what the research agenda looks like. We will mail it to everyone for your comments.

But, if you have comments before we do that, please feel free to send them directly to me or to Jim Cohn. And, we will try to incorporate them early on.

Finally, I would like to introduce our Chair for today, Paul Portney, who is currently the Vice President and Senior Fellow at RFF. Paul is a former Director of the Center for Risk Analysis and the Quality Environment Division.

He has done a variety of things. He has been a Chief Economist for the Council of Environmental Quality.

He has spent some time at Cal, Berkeley and some time at Princeton. He received his Ph.D. from Northwestern University.

And, I think, as everyone in the room knows, he has

been a very serious player and publishing very widely and participating in a variety of different events that are related to the environment and the evaluation of risk analysis.

So, Paul, I would like to thank you very much for doing this in chairing this meeting.

DR. PORTNEY: David, thank you very much. I have a few very brief introductory remarks to make.

I think before I do that I want to give both Abe Haspel of the Department of Energy and Al McGartland of EPA, the guys who put up the money for this conference, an opportunity to say a few words. So, Abe, come on up.

MR. HASPEL: Thank you, Paul. On behalf of the Department of Energy, welcome to the DOE and EPA Workshop on Contingent Valuation.

I would like to thank all of you for taking time away from your other commitments to be here today. And, now that you are here, I encourage you to share your thoughts and ideas about contingent valuation.

Your participation in the discussions will assure that this workshop is a success. And, that's what we all want.

As David noted, the purpose of this workshop is to help set an agenda for future research on contingent valuation methods. Toward this end, we have commissioned 13 papers covering a variety of topics on contingent valuations.

I hope that you've had a chance to review some of the papers before coming here today. We pushed hard to get the papers mailed to you before the conference.

My personal thanks to all of the authors who have completed their papers on time. As you all know from a variety of conferences, when the papers are still coming in it makes it much more difficult to participate. So, again, thanks to all the authors.

Another round of thanks goes to the Steering Committee that helped select the topics and authors for the workshop papers -- Paul Portney, Peter Diamond, and early in the workshop planning process, Terry Smith. All patiently discussed various topics and corresponded with all of the

authors. Without their help, we wouldn't be here today.

In addition to being on the Steering Committee, Paul, as you know, has graciously agreed to moderate the workshop. You will note that benefit from his talents as a moderator during the next day and a half.

Many of you are aware of the controversy surrounding the use of contingent valuation today. Much of this controversy relates to natural resource damage assessment regulations now being promulgated by the Department of the Interior and NOAA.

I would like to tell you a short story about a controversy that I was involved in on the same topic 15 years ago. Reid Johnson and I were both working as Economic Policy Fellows and had just joined the Department of the Interior when we were asked to participate in a project at Bryce Canyon National Park.

The Park was being threatened by the development of the old coal field about seven miles off of Yovimpa Point. And, so they had come to the Policy Office where Reid and I were

both having our fellowship and said, "Can you help us?"

So, Reid and I, being good students of this area, decided, well, what we would do, we would be real smart and we would do a contingent valuation study of the visitors and try to get a handle on what the Park was worth. Not only would we do a contingent valuation study, but we would also do a travel cost study so we would be able to see whether or not we could bracket our values.

And, we worked up this really great survey instrument. And, we gave it to the Park Service, who was going to copy it and make it available to us.

We proceeded to travel out to Bryce Canyon and arrived at the Park where the Park Superintendent had our survey. I got there the day before Reid and started going through the survey and discovered, much to my surprise, that all the contingent valuation questions had been eliminated.

(Laughter.)

MR. HASPEL: That made it a little difficult for us to do a contingent valuation study.

They didn't take out the travel cost questions, because at that point they didn't understand what we were going to do with the travel questions. But, they did understand what we were going to do with the questions regarding contingent valuation.

We had a major fight and did not win. We did not get the contingent valuation into that study.

We did proceed to use the travel costs to come up with a value of the Park and later discovered, in a meeting with Secretary Andrus and the Assistant Secretary for Fish, Wildlife and Parks, that the Park Service had taken the attitude that because they did not understand this method and because they were not sure what we were going to do with it, they were afraid that we were going to actually put a value on the Park. And, in fact, that's what we did.

(Laughter.)

MR. HASPEL: We did it through the travel cost method. Nevertheless, we received a letter the day after the decision was made on what to do with the coal field that

basically decried the fact that we hadn't even attempted to try to put a value.

Well, that was the controversy then, was even trying to put a value. I think now we have, at least, gotten past that.

And, people are looking at how do we value. It's more a question now of what method do we use rather than do we use any method at all.

And, that's I think where we are getting to the heart of the new regulations. It's how do we do it, not whether we do it.

So, therefore, during this workshop, I am going to ask personally that you all leave your personal and institutional agendas regarding the old mellow rulemaking, leave them outside the door. The purpose of this workshop is to concentrate on identifying key research issues.

From this workshop, we want to learn what type of research and studies will help you through the methods, the state-of-the-art in contingent valuation. After this

workshop is over, the Department of Energy will talk with EPA and with other interested agencies about contingent valuation research.

At this point, I cannot tell you whether where these talks will be or what specific research projects might occur as a result of them. But, I can say that the outcome of this workshop will be important in helping to direct the Department of Energy's future work in contingent valuation. I am committing to that.

Learning more about contingent valuation is important to the Department because of its potential application and economic analysis of many different energy related projects and issues. For example, contingent valuation can be used to help calculate externality values for emissions from electric utility power and can be used to estimate the economic damages caused by an oil spill.

For those and other applications, we all have an interest in getting the numbers. In the long run, the best way to get the numbers right is to develop a research

strategy aimed at improving the reliability and validity and transparency of a contingent valuation method.

I look forward to your help today and tomorrow in identifying the elements of that research strategy. As I look out into the crowd, I see faces of several colleagues who I have worked with on a number of different issues. It's sort of nice to be reading something that doesn't have the words "global warming" or "sustainability" in it.

(Laughter.)

MR. HASPEL: It's a pleasant change of pace for me to be away from the day-to-day demands of the office and to have a chance to renew old acquaintances and to make new ones. I hope that all of you enjoy and benefit from the workshop as much as I know I will.

I look forward to discussions that we will have during the workshop period and at lunch. And, hopefully a valuable and dynamic research agenda will arise.

Thank you very much.

DR. PORTNEY: Abe, thank you. Al McGartland from the

Environmental Protection Agency.

MR. McGARTLAND: I couldn't have said it better than my colleague from DOE. So, I will be a little bit briefer.

We also picked a very good day from EPA's standpoint to have this conference today. As many of you know, the west tower of EPA was completely shut down. It's sort of a delight for me that I don't have to worry about that. I will let my big senior managers worry about how to get that back into place and we can talk about real issues here.

When Howard Bruenspeck with DOE called me and asked me if we were interested in co-sponsoring this conference, I didn't have to hesitate. EPA, perhaps more than any other agency, although somebody like NOAA or DOI, has a real stake in the betterment of the contingency valuation methodology.

Not only do we use it in our regulatory support approaches and in the benefit cost analysis, et cetera, but we also want to embark on several new initiatives that we have on planning and getting some of the benefits out of that equation of the ledger right that critically depends

for many of our benefits on the use of contingent valuation.

Without sort of having an agreed upon methodology that people sort of believed in, that will certainly put a stop on our plans.

We have also used a lot of CV in support of our legislative initiatives. The Clinton Administration, in particular, is facing a very hefty legislative agenda in the environmental arena.

We have the Clean Water Act, the Sewer Fund, the Safe Drinking Water Act and others who are on track right now. And, a lot of the questions being asked are sort of what are we buying for our investments. Without a CV, it's going to be difficult to answer some of those questions.

So, I, too, and EPA, too, would like to thank all of those involved in the conference today. It's obviously a very good turnout. And, we are very pleased.

And, with that, I will turn it back over to you.

DR. PORTNEY: Abe and Al, thank you very much. Before you are all welcomed out, let me extend my welcome to you

and thank you all for being here.

A very perceptive colleague of mine caught me before I left work last night and asked if we were going to run the seating the way it's run at a wedding. I said, "Well, what do you mean by that?"

(Laughter.)

DR. PORTNEY: He said, "Are there going to be people at the door saying, `Friends of CV or opponents of CV?'"

(Laughter.)

DR. PORTNEY: As I look on the audience, I am delighted to say that, at least, for this first session it looks like a perfectly random distribution of views.

(Laughter.)

DR. PORTNEY: I'm going to be worried if there is a lot of recontracting at lunchtime.

(Laughter.)

DR. PORTNEY: Actually, I am much more optimistic than her comment would suggest. We really have an excellent program here.

I think we have got good presenters. We have got terrific respondents.

And, as I look out into the audience and look at the names of the people who have registered for this today, I mean, we really have an all-star cast out there in the audience. And, these are people that I am hoping, over the course of the next day and a half, will -- that you, yourselves, will become participants, albeit brief participants if you are at the microphone.

(Laughter.)

DR. PORTNEY: I think that we all have an opportunity here over the next day and a half to learn a lot about the research that is being done that pertains to the contingent valuation technique and its ability to produce reliable or credible or whatever kinds of estimates about lost passive use values or nonuse values. More importantly, I think it's an opportunity -- and this is an especially important opportunity from the standpoint of Abe Haspel of DOE and Al McGartland of EPA, and those from the other government

agencies here who have the ability to fund research, to learn about the kinds of research questions that need to be explored in the future.

So, I would just echo what Abe said, that to the extent that we can really keep our eye on the research ball and talk about the kinds of questions that need to be funded better, the kinds of projects that need more support, I think we are going to make a lot of progress. And, I am especially optimistic, as I stand here, because I think today there clearly is more interesting and good research going on relating to the contingent valuation question than has ever been the case in the past.

Indeed, a lot of the people who are sitting out here looking at me right now are the people that are doing that research. And, again, I want to give you an opportunity, and all of our presenters an opportunity, to talk about that and to talk about the things that need to be done.

More importantly, my experience in the contingent valuation debate has not been as a researcher. It came as

being a participant in a NOAA panel. And, that sort of introduced me to this area in a way that made me appreciate the research but also the intensity and ferocity of the debate.

One of the things that I am optimistic about today is that, at least, for the time being, the tenor of the debate seems to be somewhat less personal. And, I hope that we can keep it that way, at least for the next day and a half.

I expressed my optimism to one participant this morning. And, he said, "Well, just wait until the next court case."

(Laughter.)

DR. PORTNEY: Well, as far as I can tell, the next court case is not before us right now. What we have is, you know, an additional 36 hours or so.

Let's use that in a congenial spirit to talk about research, as I say, that is being done and that can be done. And, I think we will all come away from this better researchers and with a better appreciation of research and

policy nuances that are involved here.

Without further ado, I want to get the program going. I will provide brief introductions of each of the people before they appear here.

Once we have gotten through the paper presentations and the responses, we do have a lot of time for you to make brief comments and ask questions. And, the way to do that is go stand behind one of these two microphones.

Since I don't know everybody here, if you put your hands up I'm not confident of my ability to keep track. But, if you stand at those microphones, I will take turns calling on people from those microphones.

If you get going for more than three minutes at one of those microphones and I haven't heard a question mark, I may sort of butt in. And, I don't want to be rude about it.

I just want to warn everybody in advance. Please, be brief, because there is a hell of a lot wisdom in this room. And, I want to give everybody an opportunity to benefit from all of it.

So, without further ado, I want to introduce our first session. The paper presenter is Tony Fisher.

As most of you know, Tony is a Professor of Agricultural and Resource Economics at the University of California at Berkeley. In previous incarnations, Tony was at the University of Maryland.

And, for a brief time, I had the pleasure of being Tony's colleague at Resources for the Future. He has written articles and textbooks on natural resource and environment economics for the last 20 years.

And, he has made significant contributions. And, it's a great pleasure for me to start our program by turning it over to Tony Fisher.

MR. FISHER: Can everyone hear me? Well, let me start out and we will see.

I was asked to assess contingent valuation for consistency with economic theory, in particular construction of utility functions. And, specifically I was asked to focus on the difficulties that arise in interpreting the

results of contingent valuation to shed light on the value to be attached to changes in environmental goods. Finally, what research, if any, might resolve difficulties.

So, what I will do here is, first, very briefly, set out a framework for the theoretical consideration of the topic, indicate the role of contingent valuation and then I will indicate a couple of difficulties I see in interpreting them for use in environmental valuation studies. First, some issues raised by the hypothetical nature of the procedure will be considered, including potential for strategic behavior by respondents, the choice of CV format -- that is, closed-ended or open-ended -- and the choice of measure of value, willingness to pay or willingness to accept.

Then, I look at I, and many other people, call the embedding problem, the frequent finding that willingness to pay for a good is approximately the same as the willingness to pay for a more inclusive good; that is, one in which it is embedded.

Then, I raise the question of whether, or to what extent, externalities in consumption, such as altruistic preferences, should be taken into account in a welfare evaluation of an environmental good. The reason I got into this is through a consideration of bequest value.

As you know, contingent valuation is usually put forward as a way of getting at nonuse values. And, the nonuse values are usually considered existence value, option value and bequest value.

Existence value, we will be talking about option value, as I've argued elsewhere, is not properly a separate nonuse value but rather is something that falls out of what I think is the appropriate consideration of uncertainty in assessing use values. And, that leaves bequest value. But, I see that, as I will argue, as a subset of altruistic values.

And, I will indicate suggestions for research along the way as these seem relevant. So, let me take just a minute or two to discuss a framework for the discussion of CV.

Okay. So, the idea is to represent an individual

utility function as u(x,z), where x is a vector of market goods and z is a vector of environmental goods.

And, we assume that the individual maximizes utility by choosing among the market goods, the level of provision of the environmental goods, z, being not subject to individual control. So, the problem then is to maximize u(x,z) subject to a budget constraint, px=y, where p is a vector of prices and y is income.

The constrained optimization yields the ordinary demand functions, as you can see there, xi=hi(p,z,y), where i indexes the ith market good.

Now, we can define the indirect utility function, v(p,z,y) by substituting in for the x's in the utility function, u[h(p,z,y),z], so that indirect utility function utility is represented as a function of prices and income and also, in this case, the environmental goods.

Now, if we suppose that one element of the z vector is increased, with no decrease in other elements, and no change in prices or income, let's just suppose that the only

increase is in good z1 so that z superscript 1 is greater than z superscript 0, then the utility, that frontal view of what is called u superscript 1 equals v(p,z1,y) is greater than u superscript 0 which equals v(p,z0,y).

And, now we can get to the point of interest, which is the definition of willingness to pay, used in the CV framework. The compensation variation measure of the utility change can be represented in terms of the indirect utility function.

We have v(p,z1,y-c)=v(p,z0,y), where the compensating variation, c, is just the amount of money that, if extracted from the individual after the change in z from z0 to z1, will leave him just as well off as he was before the change. And, this compensating variation is just the willingness to pay for the change.

And, it's this amount which a CV survey attempts to elicit from a respondent. And, I should mention that since the environmental good is presumably a public good, the total willingness to pay is given by aggregating over

individuals, though it's not always clear who should be included in the aggregation, especially where nonuse values are concerned.

I don't, however, have much to say about this subject. This is the extent of the framework I wanted to present. And, I think it will be sufficient for any remarks that I would like to make.

Now, what are the difficulties? Well, one, in my view, is exhibited in a now well known study by economists, Seip and Strand, in Environmental and Resource Economics, a European journal in 1992, that compares responses to a CV survey of willingness to pay for membership in a Norwegian nature protection association with responses to a follow-up solicitation of those who indicated a willingness to pay greater than the association's membership fee. And, as it turned out, less than 10 percent of those individuals accepted the invitation to become a member and actually paid the membership fee.

There have been a couple of other studies along these

lines. Still, it's not very many.

The questions that arise have generally these results. It seems that ideally pooled findings from many studies might be used to help calibrate CV results.

The Seip and Strand study has been criticized by, among others, Stale Navrud, who is here today. One of the issues is the definition of the good.

That is, without going into a lot of detail, respondents may have been expressing values for different goods in the two phases of the Seip and Strand study. Navrud did another study in which greater care was taken to keep the definition of the good the same from one phase of the study to the other.

A couple of other careful procedures were followed. And, the results were somewhat different.

It turned out that about 30 percent of the initially positive people followed up and paid money. So, 30 percent is probably significantly different from 10 percent.

On the other hand, it's still significantly different

from 100 percent. So, it seems that a more careful definition of the good can reduce the discrepancy between stated and actual willingness to pay. It, perhaps, doesn't eliminate it.

So, how do we account for the discrepancy? Well, one possible explanation is strategic behavior, that is free riding by those asked to pay for the provision of a public good. So, rather than CV yielding an overestimate of willingness to pay, what we have is observed behavior yielding an underestimate.

But, there is a problem here, in that if individuals are behaving strategically by not paying, why should they behave strategically in stating willingness to pay? One hypothesis would be that stated willingness to pay would be an overestimate if respondents believe they can influence the outcome but will not be assessed their stated willingness to pay.

Of course, if they believe they will be assessed, the bias cuts in the other direction, as in the case of the

actual payment decisions described in the studies. Now, it's not clear to me from the relatively very few experiments that have been carried out so far how CV responses should then be calibrated to yield a true willingness to pay, if at all.

An alternative approach that I put forward is that one might try to bound true willingness to pay with recognized overestimates and underestimates. Stated willingness to pay can be taken, for this purpose, to provide an overestimate by some unknown amount.

An underestimate might be provided by an extended experiment that, in the Navrud format, say, iterates on the subscription fee to an association or to a journal to a point where the individual accepts and makes the payment. The actual payment made, in these circumstances, would lie below the initial amount which the individual refused to pay. So, the actual payment would represent an underestimate of the true willingness to pay due to free-riding behavior.

More generally, it seems to me that an experimental approach, in which people play with real money, and face real consequences of their choices, should be more widely employed. This can take the form of supplementing a CV study to yield bounds, such as suggested, can aid in calibrating CV responses or it can even serve as an alternative to CV.

When I wrote this paper, I could think of only one, or I was familiar with only one, example study of this sort, the classic by Peter Bohm back in 1972, in which people were asked to pay for a partly public good, a closed circuit TV showing of a stage performance that wouldn't appear for some months. Five different non-hypothetical approaches to eliciting payment were tried, differing in their apparent incentives to participants to overstate or understate willingness to pay.

If you recall the study, somewhat surprisingly, I think, to economists who expected to find evidence of strategic behavior, no significant differences were found

across the five approaches.

A sixth approach, in which some of the participants were asked to report hypothetical willingness to pay, was also tried. And, there has been some controversy about what the results showed.

Let me just say that in a very recent paper, Peter Bohm has taken the view that non-parametric tests that determines whether responses or results came from the same distribution as other sets of responses suggests that, in fact, the non-hypothetical approaches all seem to have come from the same distribution; whereas, the hypothetical or CV approach does not seem to come from the same distribution. So, he concludes that in this particular setting, CV produces an upward biased estimate of true willingness to pay.

After I sent off this paper, I saw some extremely interesting results in an article by -- there are a lot of them -- Neil, Cummings, Scanderton (phonetic) and Harrison -- I know Ron Cummings and Glenn Harrison are here or, at least, they were here last night --

(Laughter.)

MR. FISHER: I don't know how much I can get away with with this. What I want to do is put on the results.

It's an extremely interesting piece of work. It's called "Hypothetical Surveys in Real Economic Commitments."

In this study, three passes were made of valuation of a good. Now, not a public good but a private good.

And, the top set of numbers, generic CVM, tells the contingent valuation, what are you willing to pay for this. And, you can see mean and median responses there.

Let's look at the top row. You can see the mean is 329 and the median is 75 dollars.

Then, if you jump down to the bottom, what they did was they set up as an alternative procedure, a sealed bid auction, which the winner of the good is the person bidding the highest amount but the price that they pay is equal to the second highest bid. And, it can be shown that bidding truthfully at such an auction is a dominant strategy for the individual.

So, you can see the results. It turns out that the mean is considerably low, around 12 dollars. And, the median is about five.

I'm not a statistician or an econometrician, but this looks like a significant difference to me. Now, the author has noted that the difference could be due to the difference in the format or the provision of payment specified in the experiment.

So, they did the middle section there of the table, a hypothetical auction which was designed to test the hypothesis that -- well, one is, is it due to a difference in the provision or the format, in which case the hypothetical auction results ought to be different from the CV results but similar to the auction results. Or, is the difference between the auction and the CV due to the hypothetical nature of the CV procedure, in which case the hypothetical auction ought to be closer to the CV hypothetical and no close to the actual auction?

And, you can see the results are very close to the CV,

a mean of 301 and a median of 60. Again, I am not sure what careful statistical analysis would show.

The authors probably go through it. But, to an outsider, to me, it looks like a very close hypothesis that these results came from the same distribution as those for the CV.

So, one experiment doesn't prove a point. I think that has been stated by a number of people in this area.

Still, these are very suggestive results taken along with Bohm's. And, what I would like to see is more studies of this sort followed by careful statistical analysis, especially one that compares them; in other words, the meta-analysis of the results trying to determine what's the influence of differing valuation formats on measured willingness to pay.

I should mention also that this was an open-ended CV question, willingness to pay. It was not a referendum type of closed-ended format.

Okay. The original Bohm experiment raises another

issue -- choice between open-ended or closed-ended. This is a very lively issue, I know, in the CV literature.

And, it's worth noting that the approaches reported by Bohm as not yielding evidence of strategic behavior were what he called referendum-like. And, also, as you know, the NOAA panel recently concluded that there is no -- in this format, namely closed-ended, there is no strategic reason for the respondent to do otherwise than answer truthfully.

It's starting to get late, so I think I will just say that it seems to me that that case has not been made, either theoretically or by empirical findings. And, no doubt, there will be some discussion.

There is one other issue that I want to raise that may be related to the hypothetical nature of the CV procedure and that is the choice between willingness to pay and willingness to accept measures of value. We can easily measure willingness to accept.

As you look at the last line, this shows indirect utility function considering a change in the environmental

good from z0 to z1. Only now, the individual is not going to get it.

So, we ask, "What change in income? How much would he have to be given to make his as well off?" And, we find that k will accomplish that.

Now, it's well known that in the careful studies where both measures are calculated, willingness to accept consistently exceeds willingness to pay by as much as an order of magnitude or more. It seems that the disparity is greater than can be accounted for by income effect and even greater than can be accounted for by an income effect, as formulated in Hanemann's very interesting 1991 AER study that showed that what was at issue was really not the conventional elasticity of demand but what Randall called the price flexibility of income, which Hanemann, in turn, showed is equal to a ratio of conventional elasticity and a substitution elasticity between the public good question and some composite private good.

Now, that is if you use the Hanemann/Randall/Stahl

bounds, you get plausible estimates of these parameters. You get fairly large disparities between willingness to accept and willingness to pay.

And, I haven't looked at all the studies. But, it's my impression that still there are a number of results that you have to really strain to come up with primaries that would provide the bounds as wide as those given in some of the studies which, as I say, can be in order of magnitude or more. So, there may still be something else at work.

Hoehn and Randall have suggested that stated willingness to accept and willingness to pay values will converge. And, Don Coursey and his associates have found in recent articles that willingness to accept and willingness to pay do converge in repeated trials. And, virtually all of the movement is in willingness to accept. These are trials in the experimental format.

So, what they suggest, the thing about research agenda, is that if willingness to accept is the preferred measure, an experimental procedure involving repeated trials should

be followed to improve reliability. And, alternatively, the researcher might employ willingness to pay on the grounds that this will yield a good approximation of willingness to accept.

But, above all, it seems to me that further experiments are warranted to determine the robustness of the convergence results and, more importantly, their applicability to CV because, remember, parts of the results were in the setting of an experimental trial.

Now, even in Coursey's results, some disparity remains. I have one other argument here to make very briefly. And, that is whether willingness to accept is, in fact, the right measure, the theoretically appropriate measure, in the case of injury to a publicly-owned resource as in most natural resource damage cases.

Robert Mitchell and Richard Carson, in their volume on valuing public goods, performed a suggestion that willingness to pay might, in fact, be the generally correct measure. In the case of an ordinary private good,

willingness to accept is the appropriate measure, because the owner enjoys a generally acknowledged property right to the good, which includes the right to exchange it for money.

But, in the case of a public good, such as environmental quality, the situation with respect to the property right is different. The individual consumer of the good generally does not have an acknowledged right to transfer it, even in return for a cash payment.

And, if there is a cost to providing the good, it is borne by all consumers through some combination of higher prices, taxes and fees. So, the appropriate analogy is not payment for an ordinary private good but payment of a maintenance fee by condominium owners for common amenities, such as landscaping.

More amenity can be had for larger payments by all. Payment of a smaller fee will result in correspondingly less amenity.

So, my idea here is that we might, in a CV survey, ask for willingness to pay for each level of the good from the

initial level of income. In other words, it's true that willingness to accept can be -- is equivalent to willingness to pay from some level of income.

In fact, the initial level plus the willingness to accept is a transfusion of income. But, what I am talking about here is willingness to pay from the initial level of income.

And, you can ask how much are you willing to pay for this level of the good, how much for that level, and so on. And, you will get something like I want to show in discussion of the next issue, which is embedding.

Okay. This just brings me to -- and not very smoothly perhaps, but brings me to the next issue which is embedding.

And, granted that embedding is present in at least some CV studies, how can it be interpreted? One interpretation is just that the studies that appear to show embedding are flawed in one way or another so that the results aren't significant.

Another is that, as the results are inconsistent with

axioms of consumer theory, such as more is preferred to less, the method that yields the results is flawed. Of course, it's possible that both are correct.

At least, some findings may derive from CV surveys that fall short of the state-of-the-art; and, yet, even a survey conducted according to the best practice might well turn up anomalous findings that cast doubt on the method. I would like to explore the middle ground here.

And, I begin by accepting, at least for purposes of discussion, the finding of embedding in one form or another. Now, the question is: Can this finding be reconciled with consumer theory?

Now, that's not the only question. I mean, it's not just is it theoretically possible if we really strain to find a way to do it? Well, that's the first question.

The second, an equally important question, is: Is it plausible? I mean, do we have to really strain or does it seem like a reasonable kind of explanation?

But, we consider two environmental goods, z1 and z2.

If they are substitutes, then the marginal utility of z1 is diminished when consumption of z2 is increased.

So that willingness to pay for z1 is also diminished. And, willingness to pay for a change in both z1 and z2, taken together, will be less than the sum of willingness to pay for the changes taken separately.

And, of course, the results go through for z1 and z2 perfect substitutes or essentially the same good. This is just diminishing marginal utility for a good.

So, it seems to me that, at least, some of what has been called embedding should be expected as consistent with economic theory. But, the more difficult question is: Can substitution and diminishing marginal utility explain all of the apparent embedding? Does marginal utility really drop precipitously to zero, as it seems to do in some studies, after the first few units of consumption?

I want to go through a very simple exercise that provides a rationale for this result to occur, although ironically the example I will used is really not very

plausible or very good interpretation of the results. And, I use a study by Bill Desvousges and some of his colleagues that looks at the willingness to pay for different percentages or numbers of a bird population saved by a program to put covers on waste oil ponds that would otherwise entrap the birds.

So, the indirect utility function that yields willingness to pay for saving just one percent of the population is written as on the top line there where z1 is 1 for one percent and z0 is zero for zero percent. And, then you can do what I was talking about earlier in terms of getting willingness to pay answers for different levels of provision of environmental good if you just keep adding some percentage.

So, you have v(p,2,y-c2). Or, I could say that the c1 would represent willingness to pay or compensating variation for saving one percent. And, c2 will represent willingness to pay for saving two percent and so on all the way up to saving 100 percent, as you can see in the last line.

So, if utility is concave in z1, then willingness to pay, or c, is also concave in z1. Now, in the Desvousges study, it was found that willingness to pay for a program that saved two percent of the birds was not significantly different for a program that saved less than one percent.

However, these percentages refer to programs that get the population back up to 100 percent of what it would have been in the absence of the oil ponds. That is, they get it from 98 percent to 100 or from 99 percent to 100, respectively.

Now, how do these results compare to what we predicted from my interpretation of standard consumer theory? Well, from the curvature of the c function in Figure 1, it is evident that the distance, the vertical distance, from 98 percent to 100 percent is more than twice as great as the vertical distance from 99 percent to 100 percent.

In other words, the willingness to pay for saving two percent of the birds should be not just twice as great as the willingness to pay for saving one percent, which would

represent a proportional increase, but more than twice as great or more than proportional. Yet, it was found to be the same in each case.

Now, how can this be reconciled with theory? Well, one obvious answer that occurred to me is that respondents when perceiving the same good will say, "Well, one percent or less than one percent or two percent or a little more than two percent, that's really not so very different," perhaps in the minds of the respondents.

I think that can be a good general argument, though I want to go behind that or beyond it. You know, anyone can say these are perceived the same by consumers even though, in fact, you are specifying differences in the commodities.

And, I was wondering if there was some maybe more objective way to rationalize such a perception. And, it occurred to me that, especially for nonuse value, which is what we are talking about here, I've already said option values are not.

Bequest values I will talk about in just a bit if there

is time. But, they tend not to be that important.

And, that leaves existence value. And, what better example of existence value than existence of the species? It's literally existence value.

So that you could say that any increase in the population beyond the minimum viable, whatever that happens to be, a critical level, would add nothing to utility. So that the marginal utility, beyond the minimum viable level, would, indeed, drop to zero.

And, by the way, this is a scenario that is not restricted to endangered species. You could think a bit of wilderness areas, another example to which CV studies have also been done for.

Once protection of a viable ecosystem of a certain kind has been assured, preservation of additional acreage, or even additional areas, might be expected to add little or nothing to utility. Now, for the example of the bird population, we can represent this exercise in Figure 2 which shows the willingness to pay, or c, as a function of z1, at

zero until the population reaches the minimum viable level of z1 bar.

And, there it jumps to c bar and remains there as the population increases beyond z1 bar. So, the difference in utility between 98 and 100 is now equal to a difference between 99 and 100, just as recorded in this study. Unfortunately, both differences are zero, not the finite amount reported there.

If you look at Figure 3, now there a couple of ways in which the findings in this particular study can be reconciled with theory. But, importantly, each depends on a misperception of the good in question.

If the minimum viable population, say, were z1 bar, were just above 99 percent, then willingness to pay to go from 98 to 100 would be equal to willingness to pay to go from 99 to 100. They would both be equal to c bar, as illustrated on the figure.

I say misperception, because the true z1 bar, whatever it is, is almost certainly far below the range of 98 percent

and up, described in this figure.

Alternatively, on Figure 4, respondents could have believed that the scenarios were describing programs that would save just one to two percent of the birds. And, if they also believed that the minimum viable population was one percent, then again willingness to pay would be the same for both programs. And, that's illustrated on the figure.

But, this also represents or rests on a misperception, namely that only one percent or two percent of the birds will survive rather than 98 percent to 100 percent. A more realistic situation is probably one in which, as illustrated in Figure 5, at least some people value existence of the species and some, perhaps the same ones, value existence of individual members.

Then, the willingness to pay curve might be concave to the point of the minimum viable population, jump there to register the value of saving the species and become concave again as the survival of individual members adds some value.

So, none of this proves anything. But, the moral of

this story, in my opinion, is that definition of the good in a CV survey is of crucial importance.

To some extent, this has been addressed, I understand, by debriefing respondents after the fact of the survey. If we did that, we might learn that some or all of those who participated were making one of the mistakes identified here or some other.

But, it seems to me preferable to spend substantial amounts of time during the course of the survey checking and cross-checking the perceptions of the good being valued and not proceeding until it appears there is a general understanding and agreement on the definition, importantly including what is not included in the definition of the good.

I want to say something briefly about bequest values now. In the conservation or environmental economics literature, we often speak of bequest values.

But, it seems to me that these are properly a subset of altruistic values. It's not clear to me why a more general

term isn't widely used.

Perhaps preference for the term of bequest value represents or reflects a belief that contemporaneous altruism is empirically unimportant with respect to environmental goods. But, it appears to me that there are a couple of questions that deserve some consideration if you want to make a complete evaluation, including bequest values.

And, then there is also, I think, an answer or a resolution to a question that is appearing in some recent literature. And, I will try to indicate that, too.

The questions I have that I don't think are resolved -- they may not be very important, but this is still my 45 minutes, so I will throw them out anyway. One is, suppose individual A benefits, in his estimation, from individual B's consumption of some environmental good. Then, we would stick a term for B's consumption of the environmental good in A's utility function.

But, now suppose A also benefits, in his estimation,

from B's consumption of some other good, some private market good? Well, then consistency requires that we include a term in A's utility function for B's consumption of this other good -- call it x. There is zB and also xB in A's utility function.

But, the standard welfare evaluation of, say, a change in the price of x, the market good, involves integration of the market demand function for x, where demand does not reflect altruism or externalities in consumption generally. And, so what you may have is: Is it then appropriate to include, in a welfare evaluation altruistic values attached to the environmental good?

Another question is raised by the existence of negative externalities in consumption. A feels himself worse off when B's consumption of some item increases.

Now, it's my impression -- I don't see any discussion of this. And, I think the reason is that the economists on both sides of the contingent valuation debate are just too nice to contemplate this possibility.

(Laughter.)

MR. FISHER: I, however, can vouch for it from personal experience -- a recent divorce, which I'm not going to talk about more unless there are questions in the general, in --

(Laughter.)

MR. FISHER: -- which case I am going to have to give my side of the --

(Laughter.)

MR. FISHER: There are other examples. I mean, it's just not my former wife's uniquely look at perverse preferences.

(Laughter.)

MR. FISHER: As an example, I mention the "Nuke the Whales" bumper stickers of a few years ago. So, the question is, what do we, as welfare analysts, do about negative altruism?

If altruistic values are to be included in a welfare evaluation, clearly the sign can't matter. Yet, I suspect that most of us would be uncomfortable with the inclusion of

the negative values.

These are just some questions I throw out for consideration. Now, the one other issue about altruism that I want to address is that despite what looks like some controversy in the literature, beginning with an article by Paul Milgrom and Jerry Hausman on the critical states of contingent valuation, Milgrom essentially argued that altruism or altruistic values shouldn't be included in the benefit cost analysis.

But, what he had in mind there was altruism of the sort that says that individual A -- there is a term in individual A's utility function for individual B's utility, not for individual B's consumption of a particular good like whales or clean air or something like that -- safety.

There is a very nice article by Jones-Lee in the 1992 The Economic Journal which sorts out when do you count altruistic values and when don't you. And, essentially, it's pure altruism, the sort -- as he calls it, the sort that involves individual B's utility during individual A's

utility function when you don't count it.

And, the reason is that if you value an increment of something, say an environmental good, the level implied by people's own valuations will result in over-provision of that relative to others that they also care about. On the other hand, as Jones-Lee shows, it seems intuitively obvious, if the altruism is focused on a particular good, such as safety or environmental quality, then it's appropriate to include the full amount of the individual's willingness to pay for the other's consumption of that good, safety or quality.

And, my reading of Milgrom, for what it's worth, his response to some questions in that volume, suggested that he accepted that position and took the same position. And, that's also a point that is made in Michael Hanemann's comments here today.

So, I think that that question, unless I missed something, is generally resolved. The difficult question is when are such values significant or appropriate and what's

their magnitude?

I mean, what's the nature of the altruism? The purpose is to find out the consumption of a particular good by others and how important empirically is that. And, I think there has been very little work, virtually none that I know of, that gets at that question.

So, to sum up, it seems to me that contingent valuation faces serious challenges to its role in measuring values. These challenges involve difficulties in interpreting results of CV studies arising from both the hypothetical nature of the approach and the observed tendency of respondents to give values that reflect their willingness to pay for some more inclusive good.

To deal with the first difficulty, the hypothetical nature of the procedure, I would recommend much greater use of experimental methods such as a couple that I discussed here today, especially the article by Cummings, Harrison and their collaborators. These can bound or calibrate CV responses or can be developed as an alternative to CV, in

some cases.

Theoretical inquiry into the nature of environmental property rights in different settings may also be appropriate in determining whether willingness to accept or willingness to pay is the preferred measure. And, finally, to reduce or eliminate embedding, it seems to me that a continuing exchange between interviewer and respondent of perceptions of the good being valued is important.

Thanks.

DR. PORTNEY: Tony, thank you very much. According to the program, it says we are supposed to take a break now.

But, we are not going to take a break now. We are just getting started here.

Also, according to the program, it says that Peter Diamond has exactly zero minutes allocated to him. So, Peter, your time is up.

(Laughter.)

DR. PORTNEY: Actually, the choice I am faced with is giving Peter no time or 24 hours, depending on how one reads

that. I don't want to get a day behind and neither do I want to shorten Peter Diamond 10 minutes under any circumstances.

And, so I want to turn the floor over to Peter Diamond, who is a Professor of Economics at the Massachusetts Institute of Technology. Peter, the floor is yours.

DR. DIAMOND: Thank you, Paul. I don't know what would be worse, to have no time at all or to have to talk for 24 hours.

(Laughter.)

DR. DIAMOND: What I want to do through my remarks here is a couple of things. When I wrote the paper that was circulated to this conference, I sat down with Tony's paper, section by section, and sometimes commented on what he wrote and sometimes discussed the issues related to the issues that he discussed.

I want to tick off a couple of points there and present them here under the, no doubt incorrect, assumption that there are people in the room who haven't already read my

paper. While doing this, I was also engaged, together with Jerry Hausman in writing a paper for a forthcoming symposium in the Journal of Economic Perspectives on CV.

Paul Portney is writing the introductory paper for that survey. Michael Hanemann is writing the other paper in that survey.

It's a little bit like round up the usual suspects. So, I want to say a few words of things that are in that paper, because I resisted the enormous temptation that computers make so easy, which is to just move around the paragraphs and give the same paper twice. So, I will say a little bit about a few things in there as well.

Let me start the way my paper starts, which is picking up on something in Tony's paper, which was not in his presentation today, which is the issue of defining a nonuse value or, as it were, dividing the total value between the use value and the nonuse value. It seems to me that that's almost impossible to do.

I was interested to go back to the NOAA panel report.

And, the wording there is very careful, much more so than in some of the other writings.

They talk about nonuse value being something that leads to behavioral trail that doesn't lead to useful inferences. That's very different from the statement, "Leaves no behavioral trail."

In my written version, I won't run through the equations, but I argue that a particular mathematical formulation, which is in the literature, is not useful as a definition because it's not well defined. That is, you can divide total value between use and nonuse in a large number of different ways, all consistent with the definition.

And, in fact, you can make use value negative and nonuse value larger than total value and all consistent with the definition. That doesn't seem to me to be a useful property of a good definition.

Secondly, on the question of no behavioral trail, I started to think about the usual semantics of nonuse value and whether they might lead behavioral trails. What would

happen if somebody, on hearing about the Exxon Valdez spill, was so upset that he went out and got drunk? That's a behavioral trail.

Does that mean that that's not a nonuse value, that it's a use value? I think not.

If we have somebody who has never been to Alaska, and never going to go, the unhappiness reflected in being driven to drink, it seems to me, ought to count as part of nonuse value. So, I think this is a distinction that's hard to draw.

And, maybe my contribution to the research agenda is to suggest that everybody stop trying. This is not a partisan statement in this context because, as I go on to say in the paper, the distinction seems to me totally irrelevant.

That is, from the purposes of benefit cost analysis for the purposes of damage assessment, we are interested in total value. Dividing it between use and nonuse might interest people, might be relevant for how you combine a study of fishing with a survey that included something with

fishermen, but, by and large, it's not what we are interested in.

And, if I'm right that it's very hard or impossible to define, then it will be good not to spend effort trying, particularly if it doesn't matter.

Moving on in the questions of preferences, it seems to me one of the central questions we want to ask is what are the preferences that are relevant for a benefit cost analysis or a damage assessment. That is, what do we want to use in normative -- for normative purposes that relates to what we observe in causative behavior.

And, the simplest thing to talk about here is the warm glow which has come out of the literature, trying to explain positive behavior of charitable contributions. That literature recognizes that if preferences are only defined over what charities do with the resources, there is just no way you are going to explain observed behavior.

And, so the literature developed the concept that you get pleasure from contributing to a particular charity. I

think of it as a form of self-expression. You are willing to pay money as a form of self-expression.

And, giving to a charity or saying no in the face of somebody standing at your front door, these are things that involve forms of self-expression. And, they involve a willingness to pay associated with it.

But, it seems to me, quite clear that in the damage assessment area, warm glow should not be part of damage assessment, because the ability to express yourself by giving to charities didn't get enlarged or shrunk by a particular accident. This is really not part of what we are after for normative purposes.

And, the complication from the -- that's warm glow associated with real pains. Where it seems to me that awareness of the concept becomes important for CV is to ask the question of whether the same kind of behavior which shows up in real contributions is also showing up in answers to questions in hypothetical contributions.

And, it seems to me plausible that it is. That is, if

it shows up when your money is at stake, then it will show up as well when you are really making statements.

Now, notice here the warm glow that is involved is the warm glow associated with answering the question as part of the survey. And, that may have some connection with the warm glow that would be associated with the actual act that the survey is about.

But, that's not a one-to-one match. That is -- I've seen in the literature the statement that people don't get a warm glow from paying taxes. This has certain validity.

(Laughter.)

DR. DIAMOND: That doesn't mean that one doesn't get a warm glow from expressing a willingness to see a lot of government money spent on a purpose one supports. That, it seems to me, apart from the strategic bias issues that Tony talked about, is another legitimate concern in trying to interpret the content and numbers that come in response to the questions in a CV survey.

Now, I want to flag in this general realm of

preferences and preferences relevant for normative purposes a number of questions where it seems to me the economics literature, in general -- and I will probably throw in the philosophy literature as much as I know of it as well -- really hasn't gotten very far. These are, it seems to me, important basic questions.

One of the distinctions come in -- and a distinction some people draw and some people think shouldn't be drawn -- between preferences and values. That is, do we want to think of people as having a single set of preferences which, on the positive side, describe their behavior over all different contexts if they don't show behavior that, at least, appears anomalous across contexts?

Or, do we, instead, want to view things as people behaving in perhaps inconsistent ways across different contexts?

A second issue, not the same issue, is are we losing something in the richness of our ability to describe behavior by compressing things into a single preference

vocabulary? And, the paper which I read and found very interesting on this is Sens paper on Rational Fools, where he says when people talk about doing things that are not in their own interests, you can make sense of that statement; that that enlightens various dimensions of behavior which you would lose sight of if you said a statement like that is necessarily in error, necessarily incomplete.

So, I think drawing out the distinction between preferences and values, continuing the debate over whether that's a useful distinction or not a useful distinction and moving both in the positive arena, do people when they don't behave in a way that's completely consistent with the way they behave in other settings or do people recognize role specific behavior that varies across these segments?

And, a second question is, what should be used for policy if we see people behaving differently in these different contexts?

How do we want, we, as policy analysts, as economists advising governments, how do we want to say the normative

analysis ought to be done?

Under what circumstances should it draw off the preferences and ignore the values?

Under what circumstances should it draw off the values as being more important to the preferences?

This, it seems to me, is a rich and fertile area for considerably more thought than has gone into it so far or, at least, that I have seen. I have not read exhaustively in the literature.

Another issue that comes up within the context of preferences are what are the arguments over which the preferences are defined. This is something that Michael raises in his comment.

Are the preferences defined strictly over the state of a resource? In terms of the bird study that Tony had up on the overhead a few moments ago, the number of living birds.

Or, are the preferences defined over more dimensions such as what it was that was killing birds? What it was that was done to cut back on the killing of birds?

Are the preferences different if there are natural sources for killing birds as opposed to human-generated? People, of course, are not natural, at least when they put on neckties and stand up in front of audiences.

(Laughter.)

DR. DIAMOND: How are the preferences defined relative to answering a question?

Which of these preferences do we want to use for normative purposes? This, it seems to me, is an important question and one that touches back on the legal damage measurements in a standard way.

I've spent time, from time to time, sitting in on law school classes and dabbling in it. So, let me just, very briefly, describe what I vaguely remember from the course that I took in the 70s on the difference between compensatory and punitive damages.

Some rich nobleman goes trampling through the crops of poor peasants for the sheer fun of it. Compensatory damages are paying the poor peasants the value of the crop.

But, the law isn't willing to say, "If you are willing to pay the price of compensatory damages, go ahead and trample their crops if that pleases you." The law says, "If there is morally outrageous behavior, then it's appropriate to have punitive damages."

And, it seems to me that the question of what the preferences are defined over, whether it's just the state of the resource or the kinds of acts associated in changing the state of the resource, speak to this compensatory/punitive distinction. Unless you want to say, "No, the underlying legal vocabulary I have used is wrong," there are also preferences related to the existence value of the crops that I raised myself. And, so we should think of punitive damages as compensatory relative to that existence value.

I think these are the same sorts of issues I have been flagging here on how do we want to think of people's preferences and behavior and how we want to relate that to making policy. More generally, I think there are important and hard issues on how to design social decision mechanisms.

What's the role of benefit cost analysis and the decisions that actually come out of government? What's the role of legislatures?

Steven Brier's book on The Treatment of Health Risks seems to me a very important contribution to that, and one I enjoyed reading, on where CV fits in relative to social decision mechanism seems to me to be a useful area to explore.

I want to turn to the way my paper with Jerry Hausman starts. It seems to me very relevant to this. And, none of this is in the paper I wrote for it. And, that has to do with methodology.

One does a survey, one asks for willingness to pay for, let's say, a completely nonuse good and one gets an answer. There's the problem: How should we interpret the numbers in the answer?

We have -- CV, as commonly said, is the only way to measure nonuse values. That means we don't have anything to directly compare it with in order to answer the question, is

the survey generating the theoretical value that we want for the purpose we have.

I'm assuming the first questions I've raised have all been addressed. We know what theoretical construct we want for the particular purpose.

Are we getting back an answer to the survey? That, it seems to me, is a very tough question to answer when you don't have any other ways of measuring it.

It's hard to answer yes. And, it's hard to answer no.

One has to look at various indirect tests to do it. And, one of the things I had hoped would come out of this conference is the discussion of what sorts of indirect tests people find most enlightening, given the absence of a direct test.

One of them, which is referred to in Tony's remarks, is there are places where you can compare -- we move out of a pure nonuse good and we can compare CV for normal private goods with actual market behavior, assuming it's the actual market behavior we are looking for for our theoretical

purposes. And, we can do a comparison like that.

And, we can assume without assumptions, we will get nowhere. But, there is a carryover from the success or the way it's calibrated in one class to the other.

Similarly, we can look at CV relative to charitable contributions because, again, there is direct behavior. There is answer to surveys.

We can look at the relation of them. We can make some assumption of the extent of carryover from one to the other.

And, in this way, I think we can hope to get some beliefs that we would be willing to stand on and not be embarrassed by over issues like calibration and whether this thing has generated numbers that we can then use.

A second way is you can ask the respondents. Schultze, in a number of surveys, has asked the question, referred to in his paper here, basically: Are you embedding? I think of it with a big "R" written backwards and the u. And, a large number of people say yes.

That, it seems to me, is an awfully good question to

include in a survey. Perhaps in government regs, that question should be required in every survey.

What happens after you get the answer is a little bit harder. One interpretation is that if a large fraction of people say yes, they are embedding, then that casts doubt on the survey.

An alternate approach, the one Schultze says, is when you point out to them embedding and then you ask them to correct their answer, then you can then accept the answer that comes after the correction.

Which of these is the right response to using the question, I think, would be a good big question, a good thing to explore, to do research about, to think about. But, generally, I think that's useful.

Verbal protocols, I think, will shed light on the extent to which we think people are looking to the things that we want to measure. I come back to always saying, "What is it we want to measure?"

And, then there is the question of plausibility of

preferences. There's the question of whether if we assume -- whether it's more plausible to assume CV has measured preferences right and they are as they have shown up in the survey or, the alternative hypothesis which is, preferences are different and CV hasn't measured it.

This, it seems to me, has to be a plausibility test, because we don't have any direct tests. Indirect tests have to be plausibility.

And, that kind of split sample, parallel survey, seems to me, to be the right way of going about it.

One attraction of the adding up tests, which were the central focus of the paper on the wilderness survey that I was part of, where you ask people to evaluate A and you tell them, "We will give you A and ask you to evaluate B under that circumstance." Or, we asked them to evaluate A and B and compare the sum of the first two answers to the third, adjusted for some estimate of the income effects. And, sometimes you don't need it because there is already taxes built into the change.

The appeal to me of this is that it gets you a bit out of plausibility. If you still have the issue, "Are people responding appropriately to the survey," that question will never go away.

But, if they are responding appropriately to the survey, an adding up test, seems to me, would be more useful in the plausibility tests, because we don't get into an argument of what's plausible. Let me give you an example with, again, a Schultze study.

This is a cleanup of a Montana river. In split samples, he asks one group how much they would be willing to pay for a partial cleanup. He asks the second group how much they would be willing to pay for a total cleanup.

In order to do an adding up test, he could have had a third group to whom he describes the partial cleanup, describes the government has already committed to a partial cleanup, and asks how much -- and, therefore, it's already paid for and it's in the taxes and then asks for willingness to pay to enlarge the cleanup from partial to total. And,

now we could do an adding up test.

Schultze didn't do all three. So, that test, needless to say, was not there.

ATTENDEE: Montana didn't have the money to do the survey.

(Laughter.)

DR. DIAMOND: Okay. It sounds like you thought it would have been interesting to do if the money had been there.

I've got about a minute left. Is that right?

You are aware from my identification that I'm from Massachusetts. One of the things people in Massachusetts learned in 1988 is that if people are saying bad things about you when you work, you shouldn't totally ignore what they have to say.

So, I would like to just say a couple of very quick words on Michael Hanemann's written version -- maybe they won't be in his presentation -- where he says a number of detailed criticisms of the wilderness survey. This

conference is not supposed to be about, you know, what's exactly in the survey. But, I didn't want to just let it go by.

(Laughter.)

DR. DIAMOND: Let me say then, very quickly, two things. One is that paper was primarily about the adding on test. Things that focus attention on other aspects of it are not focusing attention on the main question. Obviously, there were bunch of questions about the survey that hold for everything.

The second thing I want to alert you to is the importance of being careful about selecting reporting data. In the papers circulated here is a picture showing the use of three of the surveys that we reported on and a regression relating willingness to pay per million acres being prevented from development.

It gets a coefficient of 59 dollars. And, it achieves a statistic of 2.2.

And, what may be the first time since I was a graduate

student, I ran a regression myself instead of having an assistant do it, because I did it last night and that was all the time I had after getting Michael's paper. I just took all seven surveys that we reported and ran the same regression.

The coefficient changes from 59 to seven. The software package I used didn't report standard errors -- didn't report T statistics but reported standard errors. The standard error was a coefficient of seven to six. So, obviously the T statistic dropped dramatically.

I think one needs to be careful and one needs a thorough examination of different studies.

I was interested, again in reading his papers, to see that there has been a meta study done examining large numbers of comparisons of CV and other methods. I think exactly gathering together as complete a set of what we have and how we want to evaluate them is what is called for.

DR. PORTNEY: And, we were doing so well, too.

(Laughter.)

DR. PORTNEY: I especially appreciate Peter's focus on researchable questions and what needs to be done. And, I'm sure that the next speaker, Michael Hanemann, will focus his remarks on that as well.

Michael Hanemann is a Professor of Agricultural and Resource Economics at the University of California at Berkeley. He has been working a lot on CV lately and has also contributed to the literature on water resource evaluation, et cetera.

Michael, the floor is yours.

DR. HANEMANN: Can you hear me? Well, I greatly appreciate the opportunity to be here and to discuss Tony's paper. I, like Peter, will diverge from my written remarks.

And, I want to talk about three questions: What do we actually know so far? What can theory tell us? And, where do we go from here?

I start with what do we know, because I was struck by the implicit approach in the excellent conference two years ago which took a very one-sided and selective approach to

the existing literature on CV, disregarding most of it and commenting on a few studies to their taste, and acting, as it were, as though they were the first white men to discover a vast continent and the issues that they were raising had never crossed the minds of lesser mortals before.

The key issue of the validity of contingent valuation and whether it would hold up in comparison to other methods occurred to Jack Lynch and Bob Davis in 1965, five years after -- four or five years after the first work on CV. And, they did a side-by-side comparison of evaluating recreation in the Maine woods and found that the two methods gave very similar answers. CV gave a bigger number, about three percent bigger.

Over the years, other comparisons have been made. And, by now, as a recent paper by Richard shows, Richard Carson, there are almost eight papers which show a side-by-side comparison of the overall preference and contingent valuation.

And, often they provide multiple concurrences. So, the

literature has over about 500 side-by-side tests.

This literature, to date, shows unambiguously a very close niche between the two methods, with, in fact, contingent valuation generally getting a slightly lower value of something like 90 percent of the estimate obtained from revealed preference. And, we also know about these issues from the voting literature, from the market research literature.

We have had market research studies that have stated preference. And, we have the literature on predicting voting intentions.

We know from the market research literature that there is a gap. And, that a fraction of the people who say they will buy within some time period don't. We also know that a fraction of the people who say they won't buy within the time period do.

And, we know that a crucial issue affecting the validity stated in market purchase intentions is the timing. That is, the evidence shows that a very large fraction of

the people who state an intention to purchase do purchase but not within the time period specified, rather within a longer period.

In my view, that's a crucial difference in voting in the choice on public goods. That is, with the private good you have flexibility to determine when the choice is made, when to buy.

With voting, the decision is set. And, you decide your behavior.

The evidence so far on voting shows two things --that voting predictions and predictions of voting intentions are very accurate. And, so record sentiment at that time.

They show that sentiment changes during the course of an election in response to changing information.

And, they also show that there is a significant portion of people who say don't know and that you can't just disregard them and treat them exactly like the rest of the sample and so drop them out. And, that if you do disregard them, you will produce an erroneous forecast.

And, indeed, this was associated with some of the major flaws of polls. But, that if you do treat them appropriately, you come up with generally very accurate forecasts.

Another thing that we know is that it's a very fragile thing in practice. And, I say this as one who has spent 20 years minding the revealed preference, doing study indicator, studying demand analysis whether for recreation or urban water demand or irrigation.

And, we note that the problem with revealed preference in practice is that you, the researcher, don't know what the choice was about. You can guess.

You can assume what the alternatives were. You can assume what the attributes were. You can assume what the prices were.

But, the problem is that in practice, those are assumptions. And, they are often wrong and explain why many empirical demand models estimated in disaggregated data have, first of all, a very poor fit and, secondly, a very

poor predictive power in practice.

Let me move on to the second issue, which is the question of what can theory say. What is the role of theory?

Economics is distinguished among the social sciences and has very strong emphasis on apriorities. And, the notion that the touchstone of validity is theory, consistency with theory, as opposed to, say, consistency with observed data.

Theory provides, in my view, general guidelines of a qualitative nature as a guide to what variables might matter and how they might matter. But, as has long been pointed out, theory doesn't prescribe quantitative factors.

Theory can't tell you that people should value a 400 page book more than twice as much as a 200 page book. Theory can't tell you that the demand for water should be at least .3 or higher.

Theory can make those predictions if you make auxiliary assumptions. If you make assumptions, you can get sharper

predictions.

But, the soundness of the theory, the soundness of the prediction, rests on the soundness of the assumptions. And, what other scientists would do, for example, with a prediction that a 400 page book should be valued more than twice as much as a 200 page book, if they find that that is consistently rejected by the data they might conclude generally that the assumption of quasi-convexity or whatever was wrong.

And, that's basically, I think, the situation here. Let me just, without going into detail, stress that one can show what factors influence willingness to pay. And, the factors that turn out to be important are -- and substitution events.

And, what matters is the substitution between not only for this particular commodity and other public goods but also between other market goods. In principle, if the substitution affects a strong, you get pronounced embedding. And, if the substitution affects a weak, you get a weak

embedding.

At that qualitative level, that's all you can say. Tony asks in his paper: Do substitution effects -- are they the only explanation of embedding or willingness to pay/willingness to accept disparities?

And, the answer is that one doesn't know. All one is shown is that they are capable of explaining this.

There may be other explanations. And, I think there are some which come from the Survey Research literature.

But, I want to stress that in my view what theory is important and useful is not the ultimate touchstone, because it ultimately doesn't bite. It doesn't make strong quantitative predictions.

And, the arguments about CV have hinged on quantitative matters. And, they come down to saying people could not have had this much about this or they could not have this little.

There is abundant evidence from many studies that willingness to pay varies from the scope of the commodity.

It's true that certain studies fell short. But, if you looked at the literature among three or four dozen studies, that would stand out as the exception, not the rule.

There is an abundance of evidence of scope, going back to Smith's early work around 1970. So, then the question is: Is there enough variation of scope?

How much more are people to value A plus B than they do A? And, this is where I come back to my point that economic theory doesn't help you there.

If people find A and B close substitutes so that they value A plus B not much more than A or B alone, that's a -- the theory, it becomes an empirical question. Is there other evidence from other sources? And, is there internal evidence from the survey or other surveys that they have as much or as little? Are the substitution effects of this big or as strong?

But, in my view, it's the empirical literature and the empirical practice which is going to be the touchstone of validity, not the -- not economic theory.

I want to make a few brief remarks which come from the second part of my paper, because I think there are methodological issues which are crucial and which will prevent any successful empirical testing or research of the sort that is being proposed by EPA and DOE. And, I think one of the key issues is the notion of procedural invariance and measurement.

Procedural invariance is an ideal. But, it's not the reality in any of the sciences. And, this has been known since the turn of the century in the philosophy of science.

The so-called Dukenpoint thesis argues that all tests of theory rest on auxiliary assumptions and auxiliary considerations. And, therefore, anything -- you don't have a pure test of a theory.

You have a pure test of a theory embedded and maintained in auxiliary assumptions. And, if those assumptions are wrong, then you are not testing the theory. You are testing something other than you intended.

Although the myth of procedural invariance is written

as in the philosophy of science, it looks around causing consternation from time to time among the unwary. I think it has played a role not only in contingent valuation, it's played a role in the history of the preference reversal literature in psychology.

And, I want to comment on that briefly. As you note, the preference reversal literature dates back to 1971. People make price gambles differently than they make -- or lotteries differently than they make choices among lotteries.

The question is: What is the interpretation of this? Does this prove preference reversal? Is this a violation of rationality?

For many years, economists and some psychologists viewed this as a phenomenon of gambles and attitudes to risk. Within the last five years, I think there is a consensus growing that it's the difference in response modes that drives the result and has nothing to do whether the items are lotteries or deterministic multi-attribute items.

There is now a similar pattern of divergents between matching the choice that has been observed in a wide variety of multi-attribute content. Why should there be a difference in response mode?

The literature offers many explanations, some of them plausible, some of them a little far-fetched. For example, it has been proposed that there is something about matching which favors certain types of attributes and something meta-physical about choice and that's why you get inconsistencies.

I think a key factor is the difference in cognitive burden. Why is there a difference in cognitive burden?

Choice involves knowing which of two points lies on a higher indifference scale. Matching involves knowing the entire indifference map, so you can trace out to the point that's -- so that you can match up the exact level of an attribute with a number of points.

Choice requires qualitative information. Matching requires quantitative information. Matching, therefore, is

more difficult.

Because it's more difficult, respondents are more uncertain and more likely to make an error. And, when they make an error, it's likely to be in a particular direction. You are likely to underestimate minimum willingness to pay and overestimate minimum willingness to accept.

There is, in fact, considerable evidence to support this in the recent psychological literature, evidence in favor of the hypothesis that it is, indeed, an incorrect adjustment required for the matching test that causes the difference; evidence that subjects have much more difficulty with matching, require much more time; evidence that the subjects that cause preference reversal are the ones that give quicker answers in matching presumably because they are dropping out of answering the question; evidence that subjects can be budged from their matching response much more than from their choice response; evidence from statistical models that if you assume, let us say, that preferences might be the same but there is a difference in

the degree of uncertainty, you generate patterns with matching versus choice. Differences in uncertainty can generate patterns of preference reversal entirely like those observed.

Let me conclude. I cite this to make a specific point. The specific point is that there is abundance evidence that by objective criteria, open-ended questions in this context, that is matching, is less reliable than closed-ended questions.

And, much of the preference reversal literature saw unexplained phenomena, because it proceeded on the assumption that there was no difference in reliability, no difference in cognitive burden. And, therefore, it couldn't explain procedure or violation of procedural invariance.

The larger point is that the way research progresses is when it finds unaccountable phenomena and probes and digs to find explanations that reconcile -- and I think that is what will happen here -- by doing experiments. For example, the addition of debriefing to protocols, that is the main

mechanism I see as establishing reliability for the method and enhancing it, finding out what is going on, what people are thinking of, how they respond to surveys.

But, also -- and I want to emphasize this -- the notion of the single set of preferences that fits everybody and that everybody can answer with the same reliability which, for example, underlies tests. Now, if you insist that people act like robots, see things like robots, give answers like robots, you don't allow for differences in cognitive burden, you don't allow for differences in preferences among individuals.

Then, methodological tests in the evaluation of individual preference, I think, will be flawed. There has got to be an openness and a willingness to model divergences among people before one compares the two methods.

DR. PORTNEY: I think the best thing to do at this point is that I want to make sure that we get a break. Let's stop right now and start at five minutes to 11, okay, 10:55. And, then we will go straight through 'til 12, since

there is not any reason to take a break at 11:45.

When we come back, we will come back -- if you will line up at the microphones, I will begin taking people in turn. And, you can direct questions to Tony, to Peter, to Michael, or you can make comments of your own.

Thank you very much. We will see you in 15 minutes.

(Whereupon, a recess is taken at 10:40 a.m., to reconvene at 11:05 a.m., this same date.)

DR. PORTNEY: I have one announcement to make. And, this is an important announcement. I want to tell you how you get your lunch.

Lunch will be served at 12 o'clock in the atrium, not the atrium room but the actual atrium to the hotel here. You turn right as you go out this door and to the left.

You will see a bunch of tables set up with teal tablecloths. And, my instructions are to say that for those people who are not artistically inclined, that means bluish-green.

(Laughter.)

DR. PORTNEY: But, it will be pretty obvious to you there.

I want to begin by giving you and the audience an opportunity to come up and make comments or address questions to Tony, Peter or Michael. The microphones are open.

Someone be the first to step up, please. Why am I not shocked that it's Glen Harris?

(Laughter.)

DR. PORTNEY: And, remember, we are going to be very brief. Glen.

MR. HARRIS: I am Glen Harris from the University of South Carolina.

I have a conceptual question that is directed to all three. There are really two questions.

The first is one that is mean versus median. It makes a difference if you take, for example, the Exxon Valdez study. Are we talking two billionths or eight billionths? These are large differences.

And, in the spirit of asking researchable questions, this is one that is certainly researchable. So, I think we know the answer to it. And, I think my comments provide the answer.

So, I would like to hear from the speakers why that is not seemingly an issue; and, if it is an issue, we would like to go over the literature and what is their answer on that?

The second question, which is a much more open one, was raised by the NOAA panel. But, I think they didn't pursue it.

And, that is the temporal aspect of damages. Particularly, in an oil spill, we can expect that there would be a very clear temporal pattern.

Initially, we will have a great deal of outrage and a high willingness to pay. Over time, as the physical damage to the environment is restored, whether it has to go to zero or not is a good question, but as it's restored that damage will reduce.

Should we be taking the integral of that damage curve or should we be going at it at one point in time? Should we be measuring present value? Should we be measuring the flow of damages for a point in time?

And, then, what does the poor old subject do? How does the subject perceive that question?

I would appreciate a response on those issues.

DR. PORTNEY: As Tony points out, that is not a subject that he addressed in his paper. Let me give both Peter and Michael an opportunity to respond.

And, which of you would like to go first? Peter.

DR. DIAMOND: I think the mean/median question relates to how the survey fits in the decision-making process. If we think of the survey as generating preferences for public goods the way we normally use public good theory and then think of decisions being based on that, then it's clear it's the means of the relevant variable.

In public good theory, you add up -- it's not the mean, it's the total. And, you get the total by multiplying the

mean by the relevant number of people. You don't get it out of the median.

If, on the other hand, you view the survey as fitting into a political process in a different way, then the median being the cutoff point on a majority vote obviously has a different role from the mean. And, I think this goes back with my earlier remarks on where does the CV survey fit into a decision-making process, either a regulatory evaluation or a damage assessment.

So, how about on the one hand -- on the other hand, is that a good answer?

DR. PORTNEY: Michael.

DR. HANEMANN: I agree. I've adopted both points of view.

I see the issue as one of a social welfare function. And, in my view, it would be appropriate to adopt different social welfare functions in different decision contexts. Particularly, I see a distinction between benefit cost analysis for government projects and damage assessment.

With regard to benefit cost assessment for government projects, my own personal view has been that the potential compensation criteria is morally repugnant. And, I find some sort of voting, whether it's the median or some super majority voting criteria, ethically superior.

I know we teach students that the potential compensation criteria is all you need to care about. And, so it's the norm in evaluation of projects. But, that doesn't make it a whit more correct.

I've argued that with compensation for damages, for natural resource or otherwise, ethically the notion of making everybody whole seems more important. And, that corresponds in that context to the mean.

So, my view is that it depends -- it requires a value judgment. And, the value judgments will likely differ depending upon the context in which the analysis is being performed. And, so in different contexts, the mean and the median can both be correct.

MR. FISHER: I can say something about the second

aspect of the question about the time dimension of damage assessment. It seems to me that if one is looking at loss use values, then the integral of the damage is the relevant measure.

If one is looking at nonuse values, then it seems to me not so relevant. And, a very important aspect of eliciting the value of some nonuse, say, existence value is to specify or, at least, provide some understanding again of what's included and what's not included, what's the definition of the good.

And, one of the substitutes would be the same good or resource at a later point -- it may be a lot later or it may be only very little later -- in time so that presumably it would make a difference in the nonuse value.

Who knows that the same resource will be available in two months or two years or three years or whatever as opposed to never?

DR. PORTNEY: Glen's question also dealt with a temporal aspect of valuation. Let me ask again my co-

panelists up here if they would like to respond.

Peter?

DR. DIAMOND: I want to disagree with what I think I heard Tony say. It seems to me even in nonuse values, temporal aspects matter a lot.

In my written paper, I give as an example of nonuse value, assume, contrary to fact, that I never went to Fenway Park and I never watched the Red Soxs on t.v. but I still cared a lot about how they did. And, I cared about how they did season after season.

Then, I've got a nonuse value. I would suggest that I would even be willing to contribute to a fund to buy a right-fielder.

It would show up in real behavior. And, insofar as you wanted to count nonuse values in case somebody came out and damaged the Red Soxs, spilled oil on their bats -- I don't know -- then it seems to me that the temporal aspect matters.

The present discounted value for the nonuse value

should matter, because you care about all of those times separately. How much you care will depend on substitution elements Tony referred to, but you care about all of them.

What I think is a different issue -- and you really touched on it by referring to outrage -- is what are the values we are trying to measure. If they include values associated with the act of damage rather than merely the state of the resource then, it seems to me, we run into serious temporal problems, both in the sense of possible double counting because each time you are asking the question you are getting an evaluation of the whole thing.

And, since the values will change over time, it's not clear to me which is the right one. That, it seems to me, to be very different from the question of the values.

This is really -- partially, it's how you define what you are trying to measure. Partially, it's in the realm of what happens when you try to measure.

DR. PORTNEY: Michael.

DR. HANEMANN: I agree with much of what has been said.

I will just add this: There is an implicit assumption that values will be very unstable over time and a few years later an oil spill -- I would like to choose one at random, the Exxon Valdez oil spill -- the values will be very different two or three years later or five years later than they were at the time.

I am not in the position to give any specifics or any information. But, I think some evidence will be forthcoming within the next several months which will surprise you -- it certainly surprised me -- about the similarity of preferences.

DR. PORTNEY: Are there questions? Scott Farrell.

MR. FARRELL: I am Scott Farrell. One of the things that I am a little worried about -- and it gets back to, in some sense, the judgment of this conference.

I am a little worried that this focus on the CV research agenda is putting us in a social science position of picking a technological winner. I thought the research issue is contingent valuation.

And, somehow -- I am not sure how to phrase this as a question, but somewhere I wonder if it's possible, in developing this research agenda, if we can identify these research issues targeted at underlying constructs and then those research agenda items that are only related to CV.

Does the panel think it's possible to develop a research agenda that identifies those sort of separate categories?

DR. PORTNEY: Responses?

DR. HANEMANN: I agree very much with the implicit assumption. I got interested in choice originally as a revealed preference modeler.

And, the one thing that was in the revealed preference work of the context dependence of preference and the need to phase that into demand models in order to do a good job of predicting what choices people will make. And, one wants to have that credibility before doing welfare evaluations.

And, my view of what has happened is that most economists, particularly theorists, have a very refined and

unreal view of preferences. For example, context is totally ignored, status quo effects, starting with people assume that you have preferences in the abstract for commodities not relative to current levels, not relative to expectations and so on.

And, my view of the debate on CV is that CV shines a very different spotlight on people, on the choices on their preferences. And, we are finding many phenomena inconsistent with notions in various utility fields.

And, so what I've argued is that it is showing us that our notions are wrong. And, that applies to revealed preference, to demand modeling, in exactly the same way as the CV.

And, so I, wholeheartedly, endorse the idea of integrated research agenda, looking at how context matters and how you pick it up, how you do a better job of modeling.

Just lastly, I argue that another problem in revealed preference in practice is that you often are interested in forecasting response to prices or attributes that are not

observed in the current data set. And, so you are extrapolating out of the range of the data.

And, that's a hazard. And, so I think it's very natural in the context of collecting revealed preference data to ask for contingent behavior data or stated preference and to look at both, not only what people are doing now in regard to the choices facing them and then to move on to suppose the choice set were different how would that change behavior.

DR. DIAMOND: I think we have already moved a ways down the road you are suggesting. All of the issues raised about the different kinds of preferences and values and how you want to do normative analysis, that's not CV specific.

And, similarly, all the issues about government decision processes and where a particular piece of information data is, again, in general. And, I think it's good to remind us that there are multiple issues.

DR. PORTNEY: Okay.

ATTENDEE: I have a question. The question is to

Michael. And, it has to do with his response to the Valdez spill that was suggested.

It seems to me that demonstrating that scope doesn't matter, that the size of the group has no effect at all, is really a very weak test and a very weak proposition. And, if we accept the standards theory, then there must be some way we add up this to be acceptable.

It has been my feeling for a long time that the research, the adversarial character of the research, or the evaluation does not do us, as a community, a great deal of credit. It may be inevitable when there is litigation pending.

At the moment, there is no litigation. There is some, but that is not what this meeting is about.

So, it would seem that we have an opportunity to think of research that would bring together people with different beliefs or different hypotheses. Many of you, I think, will get some suggestive wording for a test of the adult hypotheses.

And, I would like to ask you for your comments, whether you are in favor or whether you are against it. And, we would like to collect evidence which, in part, has to do with how to word questions and, in part, has to do with the sociology of this debate, as to whether people on the two sides of the debate could agree on a critical tests of some propositions that are quite basic to evaluations.

DR. PORTNEY: Before I turn to Michael, I want to say that, one, you may be correct, in general, Danny. If there were no litigation whatsoever, a lot of people in this room would make smaller estimated tax payments in June than I expect.

(Laughter.)

DR. PORTNEY: Let me turn the floor over to Michael.

DR. HANEMANN: I agree with the notion of jointly conducted research. I started out my career as a disciple of Howard Rather looking at multi-attribute utility and proposed going to RFF to do a research project using multi-attribute utility in the field.

The areas of similarity and difference between various measurement techniques needs to be explored. And, I am strongly in favor of that.

Let me just say, with the adding up hypothesis, depending upon the nature of preferences, you can get sub-additivity or super additivity. And, so this -- there is not a single result that the theory would predict.

You can look -- if you find sub-additivity, you can see whether there is other evidence and other experiments that would support that. But, my own feel is that utility can bend any way it wants.

ATTENDEE: In the design proposed by Peter earlier where you specify that good A is provided and then you ask how much you are willing to pay for B, given that it is provided? I mean, they are not talking sub-additivity.

They are really talking of a very specific design proposal. And, I don't see that your answer addresses that at all.

DR. HANEMANN: Well, the way to do this, I think, is by

writing out equations and functions.

DR. PORTNEY: Other questions, comments? Jim.

ATTENDEE: University of Rhode Island. I have a question that is, I guess, for Michael. But, obviously he is not going to answer.

Michael, I applaud your notion that you need to think more broadly about utility theory and not just accept that as the gospel and that the world must be consistent with that. But, once we do that, it seems to me you open up a lot of questions about how we interpret responses to CV questions and whether we can really do those in conversations.

It seems to me all that disappears once we recognize that people's behavior may not be consistent. Just as an example, in your paper, you talk about how you may not want to add up initial resources, reductions from initial resources, but instead you include those as separate arguments in the utility function.

I think you could argue equally strongly that you

shouldn't necessarily add up income and willingness to pay and willingness to accept and that those should also be separate arguments in utility function, in which case now the whole concept has disappeared. You know, somebody has paid money.

Sure, it's an increase in their income but they may also view it differently. And, similarly, if someone has to pay, but as a reduction of income they may feel somehow cheated that they have had to pay.

DR. HANEMANN: First, let me distinguish between the ways in which revealed preference or stated preference results are inconsistent with utility theory. The major problem -- one major problem with utility theory is that it specifies a very abstract function, u of x. And, it proposes no further structure.

So, for example, status quo effects the consistent with u effects. That is, the utility function is u of x and x bar being the status quo.

In other words, the theory doesn't rule out results.

It's just that it's so abstract that it doesn't preclude other more specific structures which actually drive behavior.

A second thing is that we assume that people make all choices around the global choice set at all times. And, I guess that doesn't happen.

And, I will agree that that is a violation of the theory. But, I think the theory is simply wrong.

There is the question of the global budget constraint. Do people look at their entire income when deciding everything from a vacation next year to whether to buy a 50 cents candy bar? I don't think so.

So, the issue is, given that there may be separate mental accounts, given that preferences may be context behavior, the world is more complicated. Now, what I've argued is that you had better recognize that when forecasting demand.

Otherwise, you simply produce bad demand forecasts. And, there is abundant evidence of that.

But, it's also of great importance for welfare evaluation, because there you recognize you get different answers, different welfare values, depending upon which set of preference or what was the context of the group of individuals. But, one answer to that of the orthodox economist is we are not used to -- this is too untimely, this is not what we are used to and we won't consider it.

My answer to them is that that's people. You can have a theory of fish which assumes that they don't go in water and they don't swim. It will predict their behavior badly, but it may be mathematical beautiful.

I do feel that theory has to meet reality. But, what one needs then is a set of conventions and protocols which will require agreement and consensus that we will use this set of context when measuring values for thus and such purpose. We will handle embedding by including this set of commodities.

In other words, I agree which of you have human behaviors. It means that one has to then specify how one

would go about welfare evaluation. I think that's long overdue.

ATTENDEE: A lot of the things that you have mentioned though, for instance, regret, disappointment and those kinds of things, it would be the people violating the fundamental axioms that are at the basis of existence of utility functions. And, therefore, if utility functions don't exist then the basis for -- it seems to me we need to rethink a lot of those fundamental notions.

DR. HANEMANN: Let me just say that regret violates utility laws. That's all you can say.

DR. PORTNEY: Other questions?

(Laughter.)

DR. PORTNEY: Another question, I guess I need to say.

(Laughter.)

MR. MAGNUS: Howard Magnus. I think that a very serious point that I think plagues CV which has been implicitly mentioned perhaps in discussion but I don't think has been focused on as much as it should is that I think

that almost all populations that are interviewed with CV are biased, since that in most surveys going from what cigarettes you smoke or what toothpaste you use, most people are feeling comfortable. They sort of think about it. They know the products. They know something.

I think in any population that you are going to survey, you are going to have people who may never in their life have thought about how they would feel if they couldn't go fishing or how they would feel with a polluted lake that's a thousand miles away or something like that. And, it's going to be very, very difficult to interpret just exactly what you have.

I think that that CV population is probably going to have to be extremely structured to really getting the answers.

DR. PORTNEY: I want to let my fellow panelists respond. I guess I would say that, if I may, there are decisions that people make in their everyday lives that they don't get to make all the time.

I've had recently to advise somebody whether or not to get chemotherapy and pay for it and whether it was worth both the financial and personal costs involved. That's not a repeat decision. But, that's a market decision.

But, again, Peter go ahead.

DR. DIAMOND: Let me respond to Paul.

(Laughter.)

DR. DIAMOND: Lots of market decisions are difficult decisions. When we make them, we look to other people for help.

We look to other people who have had more experience making the decision, because, say, they are doctors. Or, we look to people who we think would be good at making the decisions because they understand probabilities and statistics, which is always involved.

Where the parallel breaks down is there is a difficult social decision to be made over how much in resources to put into environmental protection. And, in contrast to the example you have just described, what we are doing, in

effect, is looking to the general public.

We are looking to people who would have trouble making this decision, whose natural response would be to turn to other people to help them make the decision. And, that's the way, as a social choice mechanism, we are making the decision.

So, the fact that there are difficult market decisions being made is genuinely a fact and I think relevant for interpreting what we do with the CV. But, I think this is where I view that, not as a defense of the legitimacy of this procedure but an example of why it may make bad social policy.

MR. FISHER: Just a remark about the example of a remote, perhaps unfamiliar or even unheard of wilderness lake site that might be the subject of a CV survey. It seems to me this is the genesis for, at least, some or maybe much of the phenomenon of embedding.

As I thought about why does this arise, why embedding, it seems to me because maybe we, as respondents, would get

confronted with the question about some remote wilderness site that we hadn't been aware of and hadn't attached any value to. But, I think the mental construct is to let this proxy form be an element of preference about wilderness or wilderness lakes generally.

ATTENDEE: I agree with all of that. And, I think that all the people in this room could tackle this problem very easy by saying: All right, there is a large segment of the population that just can't deal with decisions about chemotherapy or about taking, you know, remote situations and trying to gear them to common situations.

And, these are going to be the sorts of people who are going to be responding to these CV questions. I mean, these people just don't have those sorts of -- aren't used to making those sorts of conceptual, you know, decisions and choices.

ATTENDEE: And, they also vote for President.

(Simultaneous conversations are occurring.)

DR. PORTNEY: Okay. We've got Howard Schumann's paper

this afternoon that deals with survey responses and the confidence that we can place in them.

And, we've got papers that discuss that. I guess I would like to give other people an opportunity to ask questions.

DR. HANEMANN: Could I just respond?

DR. PORTNEY: Yes.

DR. HANEMANN: The argument about whether the greater wash should be taken seriously or not is a value judgment. And, it always amuses me that people who are happy to study demand functions and commodities purchased feel repugnance at studying their preferences.

But, let me just emphasize this: Among the -- ideally in the CV study, you would like to identify not a single set of preferences. To the extent that there are sub-groups who approach this differently, you are identifying the different preferences and perceptions of the different sub-groups.

It would be entirely inappropriate to develop a CV where you have a population that knew about this all along,

that had thought about it and the population that never knew about it and to come up with a separate analyses and estimates for each of these subgroups and allow a decision-maker or whatever the contents was to think about whom to disregard or to weight them.

That's not inconsistent with research practice. And, it's something that I personally would like to do.

There is a cost. And, this I want to emphasize. If you want to come up with a reliable estimate of some statistic to sub-population, you basically need to multiply the sample. And, basically very few economies of scale.

So, if you want an estimate for three sub-populations, you need three times the sample. That has been a constraint on many of the studies up to now.

I think if we are to do serious research, a lot of the issues then hinge on different populations and whether they behave differently. And, that means larger samples and, therefore, larger budgets.

DR. PORTNEY: Okay. Peter.

DR. DIAMOND: I would like to pick up on the remark that we let the public vote for the President, because I think that gets at the heart of the different ways of interpreting CV and how it gets used.

In the political process, we use a mix of representative democracy and direct democracy. We could, a la Switzerland, put lots more things out to general vote.

We could -- if you look across states, it varies a great deal in how difficult it is to have referendum in different places. We view the problem of making government decisions as one that's complicated.

And, we use this big mix of different ways of approaching it. Now, how does CV fit into that?

Well, it seems to me that it fits in differently if we are interpreting it differently. If CV is a measure of preferences -- and this goes with feeding it into a benefit cost analysis, then it speaks to the mean and it speaks to trying to find out something that is in preferences, a piece of information that the respondents have that other people

don't have.

On the other hand, if we think of CV as a decision device for the government, as a sample survey substitute for a real referendum, then we are making a different choice about the political process. We are saying that its role is something that calls for the median.

And, we are saying its role in the decision process is not as an input into benefit costs but as the final determination, given the way we are organizing the government decision process generally. And, I think that this whole question of what CV is relates to how it fits in.

DR. PORTNEY: Other questions or comments? Richard.

ATTENDEE: I wanted to pick up on something that Tony raises, something he thought should be an important part of the research agenda, and that is sort of the Patterson type experiments. These are experiments where essentially you set up a voluntary environmental trust fund, as in the Patterson example, and you elicit actual contributions from

one group of people and you elicit hypothetical contributions from another group of people.

In this regard, Ted Gross and I have shown that the actual treatment should always grossly underestimate willingness to pay. Whereas, the hypothetical treatment should always grossly overestimate willingness to pay.

And, that is because it's a two-step mechanism, the environmental trust fund. And, the reason for this is that answering yes to whatever amount increases the likelihood that the trust fund will become established. And, yet you should get to decide at a later point in time whether you actually get to contribute.

And, for that reason, you should sort of always generate this result. Now, maybe this is opposite from what most contingent valuation studies do.

If you say you have a referendum mechanism, you've got a basic way you can compel payment. And, that's actually what drives the difference between the discreet choice question with the environmental trust fund and the discreet

choice question with the referendum. That is the key property.

So, you should be able to generate this result with a large difference, because you've got one treatment going one way and one treatment going the other way. So, I guess that I don't see that that is sort of a beneficial sort of direction to go, given that that's not the sort of thing that most CV people are doing.

Michael, Peter, anybody?

DR. DIAMOND: I think the issue you raise, which is what is the extrapolation between hypothetical and market behavior in different settings and how it relates to the kind of question being asked, is a terribly important one. What things ought to relate to how a particular CV study answers relate to what it is we want to measure?

So, I think it's important to keep in mind that a referendum type question is still hypothetical. It's not incentive compatible unless the person answering the question acts as if it were a binding vote.

If the person answering the question thinks of it as answering a survey, then it has no incentive compatibility properties.

ATTENDEE: Actually, all you would have to really show here is that the person answering the survey thinks that the decision-makers decision to undertake the act is monotonically increasing weakly in terms of a percent yes when it's actually no.

DR. PORTNEY: Next question?

(Laughter.)

DR. PORTNEY: No, some of these -- I'm not trying to preclude discussion, but some of these things we get sent papers back and forth.

Are there other comments or questions? Howard Schumann.

DR. SCHUMANN: I felt frustrated on the NOAA panel, not having a discussion about nonuse, which I am engaged in, not the Red Soxs. I would think that if you are familiar with it, you are concerned about the Red Soxs, then that creates

a larger issue. So, that was completely new.

I guess my question is: Is there a dichotomy between use and nonuse or is it a continuum?

DR. DIAMOND: I don't know.

(Laughter.)

DR. DIAMOND: I think it underscores the basic point that I was making that it's just very hard to define nonuse, which relative to most of the purposes may not matter. But, I was thinking in terms of use and nonuse in terms of direct interactions.

I guess the parallel I was going off is in terms of externalities where the legal system recognizes your neighbors making noise as something that you can get an injunction to make them stop. But, your neighbors reading Lady Chatterley's Lover quietly rather than noisily is not something you are entitled to stop.

That the interactions that are connected physically, identifiably and physically, are what I would refer to as use. And, interactions that are mental and nonphysical are

nonuse.

Now, this may be idiosyncratic on my part, but that was my intent to make sense out of the difference.

DR. PORTNEY: If I could exercise a little moderatorial discretion here, we have three of the world's leading welfare economists on this panel. And, there was an issue that was raised in Tony's paper that both Peter and Michael spoke to in their comments that I think hasn't been ventilated very much here.

And, I guess I would like to ask each of the three panelists if they would speak briefly to that. And, that has to do with the admissibility of preferences.

I guess I would be interested in hearing what each of the panelists have to say about someone, for instance, who would be willing to pay a positive amount to see Prince William Sound or some other unique ecosystem defiled and whether or not that ought to count on a benefit cost analysis and how we ought to treat that. I think it's a very serious and very difficult issue.

And, I guess I would like, if you would briefly, to have each of you speak to that. Michael, do you want to start?

DR. HANEMANN: Let me just say that while your description is accurate for Peter and probably Tony, it's not accurate for me. But, my view of negative preferences is that they probably exist and are interesting to study.

That is, I've always been struck by the diversity of preferences. And, in the policy work that I've done where there is no damage assessment, I've always thought it much more informative, when you are giving advice, to identify the different groups.

And, if there is a distinct group that has a strong pro-feeling and a distinct group that has a strong anti-feeling, it makes sense to measure all preferences. In practice, it becomes a measure of cost.

As I say, it's a matter of identifying separate samples. And, is it worth the resources to expanding those samples so that you can identify the separate subgroups?

In principle, yes, one wants to account for the range of preferences of the different groups. And, one wants to lead you to some decisions to decide how to weight them.

MR. FISHER: Well, I think that one has to be careful to avoid double counting. That is, if the reason one values some damage to a pristine environmental resource is because, in fact, one has some positive preferences about an alternative use of the resource, this means that the site is no longer suitable for inclusion in the logging system and, therefore, we can log it and provide some jobs and some income and so on, then that's perfectly legitimate.

But, that's separate from the externality of consumption. So, you wouldn't want to double count it. You wouldn't want to include it in both the income produced by logging the site or whatever and also the -- what you might think of as a negative feeling about the environmental quality of the site.

However, it seems to me that there may well be separately negative feelings about the environmental quality

which -- I mean, in my attempt to rationalize these, I attribute it to negative externality of consumption. And, that is, I'm a benevolent guy. The prospect of Michael contemplating a serene redwood wilderness in northern California pains me, because I just don't like him very much. In fact, he's feeling good about this.

(Laughter.)

MR. FISHER: So, that's a much more -- that's separate from my feeling that these areas ought to be logged and, therefore, I value some disruption there. And, it's that second preference, the one about his consumption of the redwoods -- it's almost a nonuse value that attaches -- I don't know the answer to that.

It seems to me to be logical and to be consistent my feelings or preferences really need to be weighed into the total. And, Michael seems to be agreeing with that.

On the other hand, as I said in the paper, I still feel a little uneasy about that. The question is to what extent do ethical considerations come in and preempt not expression

of preferences but the degree to which they are allowed to count in making something like that a public choice or doing a benefit cost analysis.

There are lots of negative feelings of this sort of reverse altruism. And, very often, at least, it seems to me in a legal process, we don't allow for these. We don't count them as deserving of compensation or redress or allowing in some fundamental psychological sense they exist.

DR. PORTNEY: Peter.

DR. DIAMOND: Amarcha Sen has a paper called Dilemma of a Paradium Liberal in which he says if you want to satisfy the Parado principle -- taking preferences including all externalities as the preferences you use in defining the Parado principle -- and if, as a liberal, you think there are a set of decisions that should be left to the individual, he cites whether you sleep on your left side or your right side. And, he probably makes reference to Lady Chatterley's Lover. It has been a long time since I looked at the paper.

He shows that these axioms are inconsistent. And, that's the dilemma of the paradium liberal.

It seems to me the right way out of the dilemma was to say some externalities count and some externalities don't count. And, whether they count or don't count depends on the -- since we are talking about social evaluations, it depends on the kind of society we want to be.

I don't want to have policies based on envy and dislike. And, I have no problem saying those sorts of things don't count, whether it's people that you are disliking or owls.

DR. PORTNEY: Okay. Yes. Could you speak a little more loudly?

ATTENDEE: I have two questions. It seems like all this contingent valuation has some theory that comes from the orthodox economy theory.

But, I wonder whether there is any future research relating to, for example, a recent article in 1990 that consumers maximizing utilities might not be exactly

maximizing utility. It might be more merely optimizing a solution.

So, in this instance, maybe our assumptions on maximizing utility might not be the correct assumption. And, I wonder if there is any future research question.

And, the second question I have is for Dr. Fisher. You were saying something about results from referendum valuation results are very different.

And, one of the findings of the analysis from our research was we set different kinds of rules for referendum questions. In their answers, consumers were maximizing utility. And, I think, in contingent valuation also applied rules of how we maximized commodity to change the consumers expectations.

I wonder if there would be a future research agenda to contemplate this research?

DR. PORTNEY: Let's take them in reverse order. Tony, that is directed to you.

MR. FISHER: The long answer to that question directed

to me is I don't know. And, the short answer is yes.

I mean, it would be a good topic for future research. With respect to the properties of the particular auction in the Cummings, Harrison, et al. is something I just saw literally moments before getting on the plane to come here. I haven't made a careful study of it.

I know Glenn Harrison is in the audience. As I say, I saw Ron Cummings yesterday. Maybe one or the other of them would want to address what they did.

But, my feeling is just that this was a very interesting avenue research to pursue, one that very simply and starkly frames the contrast between hypothetical and non-hypothetical and tries to distinguish the results you get in those two settings from results that might vary because of differences in the procedure or the framing, whether it's one or another kind of auction as opposed to an auction versus a generic CV study or, you know, whatever the approach might be. It seems to me that that's -- it's only one study, but it's a very suggestive one.

I was proposing that there be more studies along those lines that would make it possible for us ultimately to do the kind of meta-analysis and say, "Well, yes, these characteristics of the study lead to these kinds of results. Those characteristics lead to a different kind and uniformly lower values or higher values, whatever the case may be."

But, I don't have, I think, anything much more specific to suggest along those lines.

DR. PORTNEY: Don Rosenthal.

MR. ROSENTHAL: Yes. I think the purpose of doing research on CVMs is to come up with a better way to do CV surveys. And, to make progress on that, we need to know when a survey is done.

The question, which is wherever there is a consensus from the panel as to when we know a survey was not done correctly so we know it's not the right one to do so we can go on to another technique.

And, we propose two situations. One is the situation Peter Diamond said in his talk where you have this adding up

test and you have widely different results.

Is there a consensus among the panel that however that survey was administered that it is not consistent with economic theory?

And, the second case would be the birds case, whether it was two thousand birds or 200 thousand birds. And, I think some of the debate is where some people really don't care about the difference between the two thousand and the 200 thousand.

So, the second question is: People, prior to administering that survey, took people aside and asked them if they cared about the difference between two thousand and 200 thousand. And, if they told them, "Yes, we care very much about them," then you got the results that were obtained.

Would that evidence constitute sufficient evidence?

DR. PORTNEY: Maybe we have got quite a high level panel and could torment them over six or seven months and ask them --

(Laughter.)

DR. PORTNEY: -- to come up with some guidelines or something so that you would know when you have a -- I don't know, just a hypothetical idea.

(Laughter.)

MR. ROSENTHAL: I would just like to know how you can make progress if there is not consensus as to when you are not doing it right?

DR. PORTNEY: Okay. Responses.

DR. HANEMANN: There is an infinity of ways of doing something wrong.

(Laughter.)

DR. HANEMANN: I mean, coming up with a good list is not easy.

DR. PORTNEY: Peter.

DR. DIAMOND: Let me say two things. First, I think it's important to recognize that split sample tests and repeated questions to the same person tend to elicit different patterns of response.

And, once you look for that, people work at being consistent when they are being asked questions with obvious inconsistencies. So, I don't know this two thousand/200 thousand question is this different that it's going to be useful.

The second thing is -- and this is something I assume we will get into in a section with Howard Schumann's paper -- surveys have different purposes. Sometimes a survey is to get a number which will be used. Sometimes a survey is to test a hypothesis.

The question one has to ask then is, do we think of the finding, which is the limited responsiveness of the scope that showed up in that survey we carry over to an alternative survey instrument and alternative mode of delivery. I think we have to recognize the purpose of the survey relative to the test.

The last thing, coming back to the big picture question, Michael has made the point -- he's absolutely right -- that without maintained assumptions, you can't test

anything. I mean, when you do the classic ideal controlled experiment and you have one thing sitting here and the identical thing sitting here, you've got to maintain the assumption that here and here that difference doesn't matter.

You can't test anything without maintaining assumptions. But, I think the question has to be that people have to stake out what maintained assumptions they are prepared to live with in evaluating a CV survey for which the numbers are to be used rather than to test the hypothesis, because if there are no maintained assumptions we are prepared to live with, then there is not only no test that a CV survey fails but there is also no test that a CV survey passes.

So, at some level, it has got to be some agreed on maintained assumptions, at least agreed on by the people who put down the regulations.

DR. PORTNEY: As Michael has pointed out, there are an infinity of ways to do these things wrong. I know there are

some of you here who think there is no way to do it right.

On that note, I think it's probably wise that we stop for lunch. It's 12 o'clock.

Remember those teal tablecloths in the atrium. Be back here in one hour.

We will start at one o'clock sharp. And, thank you.

And, please join me in thanking the panelists this morning.

(Applause.)

DR. PORTNEY: Thank you.

(Whereupon, a luncheon recess is taken at 12:00 p.m., to reconvene at 1:10 p.m., this same date.)

A F T E R N O O N S E S S I O N

(1:10 p.m.)

DR. PORTNEY: We have a very full afternoon ahead of us and I would like to get started.

As you know, the afternoon session will begin with a paper by Howard Schuman, "The Sensitivity of CV Outcomes to CV Survey Methods."

We are especially fortunate to have someone of Howard Schuman's caliber to present such a paper. He is currently a Professor of Sociology of the University of Michigan, and has for a long time been associated with the Survey Research Center there.

Howard was also a member of NOAA's Contingent Valuation Panel, and during that time became emerged in what to him must have seemed like the strange world of welfare economics.

Despite that experience, he has agreed to be with us here today to present his paper, and it is with great pleasure that I turn the floor over to Howard Schuman.

DR. SCHUMAN: Thank you, Paul. I am as he indicated, something of an outsider here. My acquaintance with economics is based on one course taken about thirty-five years ago, so I certainly am ignorant of a lot of things that are assumed here.

In addition, I didn't have any history of involvement with CV until really the NOAA Panel. The panel consisted of five economists, and I think the initial assumption was that economic theory was so powerful that it wasn't really necessary to have anyone who was involved in surveys, but there was an afterthought and I joined the panel really not because of knowledge of either economics or contingent evaluations, but because I had been involved in survey research and survey methods for some time.

And every CV estimate that I know of is based on a survey of some sort, so the estimates are subject to whatever problems are inherent in the surveys and then some, and my role really is try to enumerate some of these problems and, I hope, generate some discussion about them.

In fact, I gather I have three discussions, not just two.

(Laughter.)

I am not quite sure who I owe that to. This is an outline of my paper. I will just follow that outline.

I also, however, will just add a few points from having read the discussion papers within the last day or so. I hadn't seen them earlier. I was away. I found them quite stimulating and very thoughtful, but mostly I am going to note as I go along a few points in which -- the very few points in which I disagree with some of them.

First of all, some preliminary observations. I have three kinds. One, types of survey data. It is a common distinction within survey research between facts and attitudes, or facts and opinions for objective versus subjective types of questions. It is not a hard and fast distinction just as the distinction between use and nonuse is not completely categorical, but in general one can say that some questions deal with objective kinds of data.

Whether you ate lunch today would be considered objective if we asked you that question, because theoretically it was observable. Whether anyone watched you eat lunch or not, someone could have watched you eat lunch, so we think of that as a factual type of question.

Subjective, whether you liked your lunch or not, we really would ordinarily ask you to report. It is a more subjective kind of phenomenon.

That is one important distinction. And the second, given a set of results, the distinction between uni-variant results, that is means or percentages, as against relationships, say testing hypotheses of differences between two groups, or any other kind of relationship.

In general, empirically, questions about facts that involve relationships are most robust with survey data.

So, for example, the unemployment rate which is largely objective, though not completely so, is based on surveys asking people whether in the labor force or a number of other questions, and which if you look over time you hold

constant the method involves a relationship is going to be more robust than a question about happiness, asked where one tries to get a percentage of people who say they are happy. Because we know exactly how that is asked, the words for happiness versus contentment or some other way of phrasing it, that is attitudinal or subjective answer.

Well, contingent valuation I think tries to do something I think very difficult. It on the one hand calls for subjective kinds of questions because one is asking people how much they value a particular good, and it calls for a uni-varied estimate, the percentage of the mean or median rather than a relationship.

So, CV is attempting to do something that is very difficult, and within the survey research area we would regard as probably the most difficult sorts of things to do. It doesn't mean it can't be done, but it does mean that CV is going to be very sensitive to the problems that come up with surveys.

I also note in the paper the distinction between

reliability and validity, a distinction that was confounded with the court decision, the Ohio court, and which became the decision we had to live with on the CV Panel, which used reliability for what I think of as validity.

Reliability is simply consistency over time or over some other variable where one expects people to answer in the same way, and is an important concept but not the most basic one.

The most basic one particularly in the CV is that of validity. Whether an instrument measures what you intend it to measure. Or at least you know what it measures for sure. Between the concept and the actual operation. I think that same sort of distinction was agreed to by two members of the panel, but I think Phil Coaksley continued to use reliability for what I would call validity. It is not terribly important. It is more a terminology problem, but I think it does lead to some confusion if we don't keep them separate.

I don't think reliability is likely to be a big problem

within the valuation. That is, that one can get the same estimates if you repeat a survey, there should be a good deal of consistency over time.

On the other hand, the fundamental issues of validity are the ones that we deal with when we ask for example about the scope condition.

The final preliminary consideration, Number 3 up there, has to do with economic rationality, which I am trying to learn more about. The relevance of it to assessing contingent valuation, especially how it applies to ordinary market transactions. CV -- a lot of the questions about CV that has been raised have to do with whether it follows the axioms of rationality as I understand it.

Many social scientists, who are not economics, like myself, find it a little difficult when CV responses are said to be irrational or non-rational. Not because we especially doubt that that is the case, but because we suspect that a great many market transactions are, in fact, irrational.

We think of how we choose small items in a drug store, or large items like an automobile, and we know that we do not always act very rational.

We suspect, in fact, that the average American may be even less rational in their purchasing behavior than we are ourselves.

So, we wonder if it is fair to hold CV to standards of rationality without defining carefully what these are, and equally important -- in fact, more important -- showing that they prevail in ordinary market purchases.

And I am going to come back to that, because one of the tests that Desvousges used was whether there is converging validity between open and closed questioning, and as I tried to indicate, I don't think there would be with ordinary market purchases, so I don't see why it should occur with contingent valuation.

In any case, I gained a little insight from talking with economists about this, but I do think it would be very useful if someone, and Desvousges I think did it more than

anyone else, tried to lay out what these axioms are and show that they apply to use votes, and then ask about nonuse.

Okay, now I am going to move to a set of standard survey issues and simply talk about them one at a time under the heading of Latin II, and how they are relevant to contingent valuation surveys.

Let me say first though that I am talking about a CV survey that is actually to be used to generate a number which Exxon or some other company would have to pay. In the case of the NOAA Panel, that was the emphasis.

It is quite possible for all kinds of theoretical and experimental reasons to violate the things I am going to talk about. I don't think we want to make a fetish out of methodological requirements.

But if the CV survey is to be used to generate numbers, large numbers like billions of dollars, then I do think these points are important. That is an important distinction to make.

I have done many kinds of research where I violate

these sorts of things. But it depends on the goals of the research, and if one is producing important numbers to be used in litigation, for example, these things do become very relevant.

First of all, there is the definition of the population. It is very important theoretically, because it plays such a large role in the multiplication that produces a final dollar figure.

It is not really a survey as such. That is, the definition of the population has to occur before one does a survey, but it is obviously quite crucial.

As I understand it, for the most part it should be in terms of the trusteeship of a particular part of the environment, such as part of Alaska, and that the trustees then define the political geographic area that should be included. And I won't go on beyond that.

I think this is pretty important, and it will become relevant later. Baron's paper talked about this and some other things that is less important than many other factors,

but I think because of the importance of the multiplication that occurs, it is really quite critical, and this will come up during the day.

Second, given a population definition, a clear definition of population, one needs to use probability sampling if you wish to generalize to the population. It is true that experimental differences between the conditions will often show up fairly well in a rough way with convenient sampling with one kind or another. But this can't we guaranteed.

Interactions with selected sampling factors can occur, and the variation can be due to a part of the population that doesn't happen to show up in malls, and therefore if again you are trying to generate a final number, one does need to use probability sampling from a well-defined population, and we will see why, in a very interesting finding that was produced by Desvousges, it becomes problematic partly because of the lack of probability sampling.

Third, non-response. Surveys -- it is very hard to evaluate the quality of the survey, and one of the few things available -- unfortunately, really -- is the non-response rate, and sometimes too much emphasis is placed on that. But that is the proportion of the sample that actually agrees to be interviewed and completes the interview. It is about the only quantitative measure one has of the quality of the survey. It certainly doesn't tell you everything about the quality. It just happens to be something that produces a number.

And I think in the most recent NOAA regulations, seventy percent response rate was stated as a minimum, and that is probably reasonable. It used to be eighty or eighty-five percent. What happens is as the percentage of the population that is willing to take part in surveys goes down, we adjust what should be the minimum to fit reality. And somewhere between seventy/seventy-five, maybe occasionally eighty percent, is about the best that can be done with most national surveys.

Fourth, Mode of Administration, of which there are three main forms; face to face, telephone and mail.

There has been a lot of controversy in this area. It is a little beyond me as to why Schultze included Don Dillman's letter, which I had seen some months ago, defending mail surveys. Don Dillman is a real expert in this, contributed very importantly to the success or use of mail surveys, and I have followed Dillman's approach, most recently actually, in making a comparison of mail surveys with face to face surveys. So, I have a lot of respect for him.

But the letter he wrote is not really a very persuasive one. There are an awful lot of reasons why mail surveys taken alone -- if, again, you are interested in generating an important final number rather than doing some sort of interesting experiment along the way, why a mail survey is not adequate for a contingent valuation estimate.

It has one enormous advantage. It costs a lot less than any other way to do a survey, because you basically

just need to pay people to put things in envelopes and put stamps on them and send them out.

But, despite that, there is no accessible national listing of households to create a sampling frame for use with probability sampling. Sometimes telephone directories are used. It is very hard to know the percentage of people who don't end up in telephone directories or end up with wrong addresses, and Don's letter questions something in the NOAA Report which probably was too high, I think he said, which misses about half the population. It is probably more like forty percent of the population, but that is a lot of people to miss if what you are trying to do is get an estimate of a population value, and that is forty percent who are not in the sample to begin with, plus the non-respondents.

So, you are really dealing with almost certainly a very biased kind of a sample.

The non-response, as I indicated, is certainly non-random. I made a comparison of mail and face to face

surveys. You tend to lose the less educated, tend to greatly under-represent the black population of the country. There are a lot of selective factors both in non-coverage that is not appearing in telephone directories and non-response. It is far from random.

There are big problems of functional illiteracy with the American population reading a questionnaire, understanding it, actually just having the patience to read through it. I know myself, because I get mail questionnaires all the time, and it usually will say this will only take you fifteen minutes, and I usually finish it in about three or four, which means I haven't done a very careful job in really reading the questions.

You can look with CV surveys, usually the idea is to present the scenario to do it carefully, in a certain sequence, and then finally near the end ask the valuation question.

With the mail questionnaire, obviously people can thumb through it. They can look at the end. They can look at the

beginning, so you have no control whatever over how the person is considering the questions.

There are probably some other reasons, too. There are just a whole host of reasons why the mail approach is not a good one if you want a serious number.

Dillman in his letter noted that the NOAA Panel ignored one big advantage of mail surveys. They do give a greater sense of anonymity. And, indeed, I think that is true. I have recently done a study, and we get evidence that racial attitudes appear differently to a mail questionnaire than to face to face. They turn out to be less liberal when it is done by mail, and we interpret that to mean a greater sense of anonymity, or willingness to express maybe their views that are counter-normative. And I think that is true.

Dillman, however, hadn't really read the report carefully, because the report does deal with the effects of people knowing that someone is listening to what they say, and suggested some ways to test that. We will come back to that in a moment.

Self-administration can occur with mall samples or any convenient sample. I think the big problem there is whether people take this seriously.

As I said, when I get a mail questionnaire, or if I were asked in a mall to fill out a questionnaire, I would probably do it as quickly as I could, and get on with why I was there.

So, I do think there are real reasons to be concerned about the seriousness with which people handle most self-administration. Perhaps, paying them a lot, although we don't know the effects of that on their answers, is one approach.

At the very least, for an important final number one should be able to show that the self-administration produces results that are comparable to those of careful face to face interviewing, unless anyone can explain the difference in a way that favors self-administration.

The distinction between face to face and telephone interviewing is a much more difficult one. There really are

not important differences in results so far as I know between -- actually, when the telephone first came into use, it was very easy to hypothesize all kinds of important differences between purely audio sort of thing, you can't see one another, and face to face interviewing. A whole set of comparisons failed to show important differences between those two.

It is probable that people pay more attention if you are sitting in the living room, and they can't hang up the phone, than to a telephone interview. And, of course, with telephone interviews one can't use anything graphic unless you managed to send it in advance.

But I think there the difference is a real trade off. The telephone is cheaper, face to face probably gets more concentration, and you can use pictures and so forth.

Let me move on to interviewer effects.

We do know that there are important interviewer effects, both on the telephone and face to face. They show up particularly strongly when the interviewer can be clearly

identified with the question. So, black interviewers get very different results from white interviewers if you ask questions about racial issues. They don't get different results if you ask questions that have nothing to do with race.

Same thing occurs with gender. Men and women as interviewers will produce different answers, again on a particular question. Whether this will occur with CV, I don't know.

We do know from the Alaska study that most people, or the majority of people think of themselves as environmentalists, and they may -- and therefore respondents may assume an interviewer coming to see them is pro-environmental, whatever that may mean, and try to shape answers in that direction.

The NOAA Report suggested two experiments. Let me just give the more extreme one, where the interviewer presents the whole scenario, even presents the question, perhaps, but doesn't take down an answer. Instead, gives the respondent

something to write down the answer on, give them an envelope, and lets them mail it back to headquarters without any identification whatsoever.

In other words, to try and create as much anonymity as possible. Compare that with a standard CV face to face interview, and see if it makes a difference. I don't know whether it would or not, or whether the difference would be large enough to be important.

But it is certainly something to be concerned about.

Let me turn to questions and questionnaires, which really gets more into my own interests, and I must say that one thing I did do was use the questionnaire for the Alaska study, and interviewed half a dozen people with it and I also used the Desvousges questionnaire, even though it is meant to be self-administered and interviewed some people with it, and I strongly recommend that one tries to do that, because hearing yourself ask questions and watching how people respond is one of the best ways to get a sense of what is really going on with CV studies.

There are actually two parts, as you know, to CV questionnaire. First, is the scenario, or it is to a full CV study, that describes what is going on. It is obviously important that this be done in a neutral way, and that is no doubt difficult.

I showed the Alaska questionnaire to my class, the pictures -- whenever we came to a picture of an otter -- I don't know how many of you have seen this, but there is a picture of an otter in that class. People say, "Oh, that is cute," and so forth. And pictures can have big effects.

So, I was worried about using pictures unless one is tested through split samples in some way. The effects of particular pictures which go beyond any words that one had used in the questionnaire.

You also need, I think, to check on whether people have really learned and accepted the scenario. The Alaska survey, most people indicate that the damage done was less than the media seemed to present. Everyone I interviewed, using that survey, said that. But not everyone really

believed it.

A couple of people afterwards, I asked them, and they really did not accept that the damage was as small as the survey had kind of carefully presented. It was not just what you say, but how much people have absorbed it, and how much they accepted it, and it seems to me very important to legitimize four respondents at the end of the questionnaire, their willingness to tell you whether they really believed what you were telling them.

Just because you said it doesn't mean that they have accepted it, and they have answered in terms of it.

Let me move from the scenario to the questions themselves. Again, there is a lot of controversy over this, and I find it hard myself to understand it.

It seems to me that the kinds of questions should be closed questions that present a particular price. Not an open question that asks people to simple come up with the value.

I tend to call these referendum questions, but I think

that is a little unfortunate. These are posted prices. In recent, in the last two to three weeks, I had the experience of buying a book, buying a golden retriever puppy, and trying to buy a house.

Now, in every one of those cases a price was presented to me. In the case of the book, I had no choice. I either accepted the price or I didn't. Nobody said, "How much would you like to pay for this particular book." If I had, I probably would have said ten cents, fifteen cents. In fact, the book was thirty dollars.

This is why I can't imagine why one would expect an open question and a closed question to yield the same result. This doesn't make any sense to me.

The golden retriever puppy cost four hundred dollars. You could try to find a cheaper one, but there are certain things one is looking for, and you either accepted it or you didn't.

With a house price, the price is usually not the final price. One can offer five, ten, maybe fifteen percent less,

but you don't offer fifty percent less in most cases if you hope to buy the house.

So, I just can't see why one would expect the two forms of the question to yield the same results. Moreover, when I ask people the open question about how much would you pay to save so many birds, they looked at me very blankly. And these were intelligent, educated people. They had no idea how to put a price on that.

So, I just cannot see -- if the analogy is to use goods, why one would use open question. In addition to that, to the extent that the model is a referendum model, not instead of, but in addition to that, you are using a referendum model, again, it is always phrased as a closed form of question.

I am not saying, by the way, that there isn't anchoring with closed questions. Obviously there is anchoring. If someone offers you a house for a hundred and sixty thousand dollars, you are anchored at a hundred and sixty thousand dollars, and you start to go down a little bit from that.

That is what pricing is, to anchor people. But it is not an artifact of CV. It is ordinary purchasing.

I think open questions are likely to produce mostly a mixture of nonsense and strategic answers. If you are asked to put the price yourself for something you want, you obviously should make it lower, and it seems to me no surprise the comparisons I have seen, I think they show that the open is a little lower. I would expect with use purchases it would be a lot lower than the posted price.

I do think, however, that with closed questions, it is called a referendum question, it is important to include something that I don't believe was in the Alaska study. That is, the option for people to say, "Don't know."

Lots of survey questions, people do not have an answer. I think when I asked Peter Diamond about the use, nonuse distinctions, his first response was, "I don't know." Since people are being presented with something they often know very little about, and they may not want to give an answer, but they feel pressure, but we do know in surveys that

people almost always try to answer questions if they are asked, no matter how ridiculous the question may be.

It is crucial to allow people to opt out, and to do it in a way that protects their self esteem.

So they don't have to admit ignorance, and say things like, "Do you prefer not to vote on that particular issued?" To make it very easy for someone who doesn't want to give a value to do so.

Okay, let me move to the last one. At the end of the paper I try to deal with fundamental issues of validity, which are, in the end, the most important, and just mention certain problems.

The first is that we have fairly good reason to think that in many surveys people give what are conveniently called, "non-attitudes." That is, they give some sort of answer which though -- they don't realize it is a chance answer, because we can always explain and rationalize whatever we say, nevertheless, looks as though essentially it is much like flipping a coin.

There is the possibility that in CV studies when one is presenting people questions about things they don't have experience with, and making it fairly easy to answer, however -- do you accept or do you agree or not agree to a particular price, that they will produce something like a non-attitude.

It is something very hard to study, but one of the best and most useful things here is simply to do test, retest, reliability studies over time, to see if people are consistent from one point in time to another. For CV, you ordinarily want the time difference to be large enough so they are not simply remembering what they said the earlier time.

Either they have an intrinsic value, or they construct one at the time, but they do it in the same way, and it is not just a test of their memory of what they said earlier.

If they are not answering in a chance way, then you should get pretty hefty reliabilities over time. In the case of CV, the values themselves should look pretty stable

over time.

Second fundamental issue is usually referred to as the attitude behavior problem, a long one in survey research, and it has come up in the form earlier of hypothetical versus actual purchases.

That is, will people do actually what they say they will do. There is no real way to do this in CV, because there is no real criterion here. The only thing that I have been able to think of, which is not terribly useful, is to apply CV studies to known groups.

So, one goes to an environmental group, one goes to a controlled, economic development group. They ought to produce very different values.

So, known groups validation at least will get at some of the kind of validity study that is useful. But it won't handle the question of whether a precise value is the correct value. But only whether the differences occur where one expects them to occur.

One can treat the CV study as a direct simulation of

economic behavior, and that seems to me where the scope requirement is important if, in fact, scope conditions apply to ordinary purchases. And during lunch I was talking with someone, and it is not so easy to say exactly what one should expect in terms -- if you change the quantity of something, how much. You should find the linear relation between that and the dollars people would provide.

This is, again, where it seems to me we need to see what happens with ordinary goods in ordinary markets before we demand something of nonuse goods.

Finally, my own favorite approach, though it is not one that I guess appeals much to economics, is to think of CV as a kind of referendum. Because we really do decide things in this country with referenda, often very important things. Referenda involves real decisions by citizens, often about issues that they have only a remote stake in. Very close to nonuse. For me, with children who are grown up, many local issues are almost of a nonuse nature. More money should go for schools, or for parks, and I vote in terms of my values

largely rather than in terms of interest in the traditional sense.

Moreover, the criticisms of CV that -- the CV questionnaire has a number of artifacts. For example, the order of questions affects results. Those are all legitimate criticisms, but they also apply to real referenda.

You know there are arguments over how to word a referendum question, political arguments, because how it would be worded would effect how it will come out. The order of referenda, the number of referenda in a particular election. All those have effects but, nevertheless, we take referenda seriously, and one can treat the CV survey as a real referendum, or hypothetical referendum, I guess.

Try to make it as real as possible. And as we try to make it as real as possible, we have to realize that the real referendum first of all exists in time, as was mentioned I think this morning. And second, it is always carried out in the context of political debate if it is of

any importance at all. And thus, if a CV survey is to be treated as a real referendum, and I realize this would be very difficult, it should try to simulate a campaign and make the time for voting during the survey itself as the actual time of voting.

The scenario should include in a forceful way the arguments on both sides of the issue. Those opposed to a particular willingness to pay value, say would be in the form of a tax if it is a referendum, would certainly bring up just the kinds of arguments that came up in the public hearing that the NOAA Panel held.

One would expect the opponents of CV to stress the total amount of societal money to be invested in this single prevention.

A large number of other environmental and human problems that should compete with that investment. The possibility of a solution proposed would cost more, as they usually do, or do less than is claimed. All the kinds of arguments that occurred in an ordinary referendum campaign.

The proponents of CV would no doubt emphasize the seriousness of the problem that was being addressed. They would probably show ads of the otters dying, covered with oil, in order to elicit sympathy. Emphasize the tiny cost per person. Probably the cost of this particular environmental method is no more than people have spent on bubble gum per capita over a year.

In other words, there would be a real political campaign contained within the scenario.

The Alaska survey has a tiny bit of this, and reminds people just a bit of other things that they might spend money on, but it is one of the early part of the questionnaire, and I imagine most of it is forgotten and not taken too seriously when the focus is suddenly -- not suddenly, but increasingly on the damage done in Alaska.

In the end, the respondent would need to decide among these arguments, just as in real referendum. I don't claim it would be easy to develop such a scenario. It might even be impossible, but if the idea is that CV referendum

questions are intended to simulate real referenda, I don't see how such an approach can be avoided.

Let me talk about one final issue, which occurred to me at the first draft of my paper, as I reread chapters in my paper as I reread chapters in Desvousges, which I found quite interesting.

The fact that he is discussing it is just a coincidence. His study -- not the bird study, but the second study in this monograph, which compared open and closed questions, made something very salient to me, which I earlier missed, but it also showed the limitations of the study trying to generalize beyond.

The study had a broader purpose. It involved the self-administered -- part of it involved the self-administered referendum-type question about reducing damage from oil spills in the United States. Respondents were asked whether they would pay a certain sum each year, not just one time but each year, in increased petroleum prices, because of the cost.

Different referendum values were asked to each of six randomized samples. Six separate samples, randomly created from a single larger sample of unknown origin.

And the one that is of the most interest to me was the highest dollar amount proposed. A thousand dollars a year to address this problem. The oil spills were not described in horrendous terms, yet despite that approximately a third of all members of one of these samples said they would make a thousand dollar payment every year and increase the oil prices in order to address the problem. Now, frankly, I found that completely implausible. I couldn't imagine that any of the people -- maybe two or three millionaires who spent that money actually. And that is why I find it hard to take seriously such a payment agreed to in the survey context.

It is interesting to note that the proportion of the third is about the same as has been obtained in experiments that I have done and others have done with attitude surveys, where Americans are asked to give their opinion about

fictional objects that they really don't know anything about.

But as I say, most people answer a question in a survey, and you can ask them something about something that doesn't exist, or that they have never heard of, and about a third of the population will say they are either for or against it.

So, it is about the same proportion.

Furthermore, by the way, it is about the same proportion that said yes, I think, in the Alaska study to the highest value, although it was not an implausible value.

Well, although this finding might seem limited to CV surveys where implausible values are offered, Daniel Conoman has pointed out to me the possibility that since this is a randomized sub-sample, the same or equivalent people may well be hidden within the proportions that say yes to much smaller values.

So, even if you offer only ten dollars as a value, I think the Desvousges study, like seventy percent said yes to

that and, of course, that is much more plausible.

Nevertheless, that seventy percent would include a third who would have said yes to a thousand dollars per year. And that is a very troubling thought, that even when one offers very plausible values, there is a substantial proportion of people who would have agreed to something totally implausible and, therefore, we really can't take their answer very seriously.

If, indeed, a third of every sample, whatever the value offered, no matter how astronomical, is ready to agree with it, CV surveys may be including a great deal of misleading information.

As I said before, it is interesting that the Alaska study found just about a third of respondents agreeing to its highest value condition, a bid of a hundred and twenty dollars on a one time basis, which does not seem on its face to be implausible.

Well, so the question is, something like a third of the population willing to say yes to almost any bid, that is

something we cannot infer from the Desvousges study, and this goes back to the requirement of defining a population carefully and drawing random sampling, because this was a convenient sample done in a mall.

Although it was strong on what is called internal validity, that is, the comparison of findings from one sub-standard to another, it is extremely weak in terms of external validity. The ability to generalize beyond the experiment itself.

First, the mall sample was not randomly drawn from the U. S. population, nor indeed from any definable population, and there are no reports of non-responses, I recall, in the monograph. So, we really have no idea what that overall sample represents, and that third -- if that had been done with probability sample, U. S. population, it might drop to less than one percent. Conceivably, it could go up higher, but I doubt it.

Second, the scenario was not developed in the detailed way that was true of the Alaska study, and while we cannot

be sure that this makes a difference, we also cannot be sure that it does not make a difference.

Third, stopping people in malls to fill out a brief questionnaire may not produce the kind of thoughtful consideration that proponents of CV claim is important.

Hurried shoppers may be tempted to say yes by demand characteristics of the questionnaire, ambivalent by their reasonable desire to get on with it, the purpose that brought them to the mall in the first place, which was not to fill out a questionnaire.

In sum, although the one third figure obtained by Desvousges and his associates is quite provocative. At this point, we cannot be at all sure of its meaning.

At the same time, if we turn to the Alaska study as one carried out with probability sampling, and very careful questionnaire development and administration, we are unable to estimate the proportion of implausible answers because the highest value offered to respondents was not outside the realm of probability. Thus, there is no way to get a handle

on this issue at present.

Tom and I are talking about possibly trying to do something that would combine and address this when it occurs.

At least the proposal for research that will create both internal and external validity, for that, although it is difficult to do, one does need to do good sampling from the general population and to be careful to include along with a set of plausible dollar amounts, something that is wildly implausible to see how many people will agree to it.

And then we can try to estimate the proportion of the American population, and we could try to correct for that also. I don't see this as just a test of whether the CV is worthwhile or not, but as a way of trying to adjust the results.

Let me stop here. I believe there is still a lot of further research that can be done that will be useful. I agree with the Tom, again, I think who said that it would be valuable now if proponents and opponents and maybe some

neutral people were involved in the design of the study so that it didn't become too much polemical. I think I interpreted that to mean that grants would only be given to groups that have representation from all sides.

I didn't mean that. I just mean that some collaboration would be useful. I will stop there.

Thanks.

DR. PORTNEY: We killed the feedback. Did they do that by turning off the microphone? Can you hear, Betsy?

ATTENDEE: No.

DR. PORTNEY: Okay, good. Well, I will tell you what was said afterwards.

(Laughter.)

We have three discussants for Howard's paper, the first of whom is Bill Schulze from the Department of Economics at the University of Colorado. On route, I understand, to Cornell University.

In order to get back on track, each of the three discussants has generously agreed to restrict his comments

to fifteen minutes rather than the twenty minutes that have been allotted.

I turn the floor over to Bill Schulze.

DR. SCHULZE: Paul did ask us to --

DR. PORTNEY: Two did graciously say they would do it, Bill said, "No way."

(Laughter.)

DR. SCHULZE: Okay. The question I am going to try to answer just in part, otherwise I would need the entire twenty-four hours that was proposed earlier, is how do survey design decisions affect contingent valuation, and I am going to argue that survey design decisions have a large effect on the outcome of these surveys which argues that we need a lot more research in the end.

But some examples of the kinds of decision -- can people in the back hear me?

DR. PORTNEY: Here, try this. Our reporter has to --

DR. SCHULZE: I teach classes of four hundred students without a microphone.

As to samples of the kind of design samples that have to be made, those that were made in the Alaska study, and most of these have been partially endorsed by NOAA in the regulations, the first of these is the use of a referendum approach as opposed to the use of questions which ask for an open-ended willingness to pay.

It turns out that there is actually quite a literature that one can refer to, so I am going to show you eight or nine published articles that really look at -- first are open-ended willingness to pay questions biased as compared to laboratory auction bids using real money.

There is substantial literature there, and Tony Fisher suggested that we needed to do more experiments like that. It is a perfect median, because actually there are quite a few experiments that have been done in addition to the ones described by Ronald Cummings, that we will hear about later on.

And then I am going to look at studies that have compared dichotomous choice, which is very similar to the

referendum approach, a yes/no response to a posted price, and look at what the difference is between answers to open-ended willingness to pay questions and dichotomous choices, and that allows us to get some idea as to whether or not answers to these questions are biased or not.

A second issue is design. Design is in the Alaska survey. To simply try and control embedding by providing a lot of contest. So, the Alaska study, no debriefing question was asked, which says, did you embed? Do you really want to assign all of these values to prevent another Exxon Valdez accident. There is no dis-embedding question.

I simply don't have time to get into embedding, okay, but I just wanted to raise that as an issue. It would have a big impact on the size of the values you were going to estimate.

Another design decision was the use of a lump sum bid rather than an annual bid. In other words, this is like asking you how much would you be willing to pay as a lump sum for all the milk you are going to drink over the rest of

your lifetime? That is a lump sum bid, as opposed to asking how much would you pay per month for your milk?

So, that was a decision designed to see how the money would be collected. I am not going to discuss that either because of the lack of time.

I just wanted to indicate as you can imagine, that is an interesting research question: annual bid, some bids, monthly bids and answers might be quite different, depending on what question is asked.

I think there is a question of survey administration. In person interviews, or do we use something much cheaper, like a mail survey, and can we really demonstrate a significant difference between those two response modes. I will discuss that point, and now I want to suggest that there may be other approaches which may be merely excluded by the current NOAA regulations, if you read them carefully, that we might decide pretty soon that we like better.

I am not sure the market research approach -- if I have time, which I will talk about briefly -- is a better

methodology. I think it is an open question, but presenting people a whole variety of choices which include apparently a lot of substitutes, and I would argue to get people thinking about substitutes when they answer valuation questions, might really be a better approach than the more traditional CV approaches, and if that is the case you want to make sure the regulations don't really exclude other approaches for trying to elicit values.

So, I am going to start off talking about what evidence there is in terms of methods for eliciting value.

The first laboratory economic experiment that I am aware of is one that was conducted and mentioned today already by Don Coursey, John Hovis and myself. It appeared in the 1987 Quarterly Journal of Economics.

The commodity used in that experiment was sucrose octa acetate. It sounds nice. This is an extremely bitter tasting liquid, which is non-toxic. What we had the subjects to do was to give us their willingness to pay not to hold the dental cup of this stuff in their mouth for

twenty second before spitting it out.

Now, we wanted an unusual commodity for which people had not had market experience, okay? So, did we succeed.

I think this is an unusual commodity. What we did was we first asked for a hypothetical value, after carefully trying to describe what this experience is like, and it is an unpleasant experience, let me assure you.

We gave them a monetary balance. We told them they could do whatever they wanted with that money. Fully gave them an opportunity to buy their way out of that experience. We used a victory auction, and they submitted willingness to pay in the form of bids to buy their way out of this experience, okay? So, then we could compare their initial hypothetical bids based on a description of the commodity to their actual bids to avoid the taste experience.

In the middle, they got one drop of this stuff put on their tongue so they could see what it was really like, okay?

So, the ratio of the hypothetical bid to the actual

money auction bid, the actual money auction bid, was .7. In other words, we had a thirty percent underestimate of willingness to pay to avoid tasting this horrible stuff. Believe me, you want to spit it out as soon as you put it in your mouth.

So, here is evidence that hypothetical values at least are not wild overestimates for this particularly unusual situation.

A second laboratory experiment which, again, compares hypothetical values to real money auction values is what I call the tree experiment.

This was supported by the Forest Service, and the first experiment was supported by EPA.

The commodity in this second experiment, which occurred in the American Economic Review in December 1992, was a Norfolk Island pine tree that -- it is a tree about this tall, with very beautiful leaves, a quite attractive house plant, and the subjects sat in their corral and sitting in their corral was this sweet little Norfolk Island Pine. The

subjects were not students, but staff members of the University of Colorado. So, they actually had a home to take it to at the end of the semester.

And we asked them first, hypothetically how much they would pay to purchase this tree, essentially as a house plant, to take home. Just for your information, the mean bid was about five dollars, and there was about a fifty percent overestimate in the hypothetical bid.

So, they were saying about seven-fifty hypothetically. After they went through the auction mechanism, which in this case was the Market auction mechanism, and they knew they had to pay for it, the mean was about seven fifty as opposed to five dollars, so we had a fifty percent overestimate in terms of willingness to pay.

I should point out that the hypothetical values were very unreliable as compared to the actual values. So, it is very hard to get reliable estimates, but I am only dealing with willingness to pay.

The second, we told people if you don't buy the tree,

Bill Schulze is going to take that tree in the back of the laboratory and he is going to kill it by cutting it off at the base. I am going to kill the tree. Hence the name willingness to pay to kill the tree. But you don't have to see it, because we were trying to simulate the nonuse value.

You don't have to watch it. We will randomly draw a witness who is willing to see the massacre of the trees.

So, we -- I am not claiming we captured nonuse. I am trying to get the element of concern for the life of the tree, even if you don't buy the tree. We got substantially larger values in the cases of willingness to pay, of course, if it is treated as a house plant, but we got vastly largely values in the case of willingness to accept. People seemed to take responsibility of killing the tree was theirs if you are buying the tree from a firm. If they sold us the tree, they felt responsible for the tree, less so in willingness to pay, since they never owned the tree, and if Bill Schulze wanted to kill his own trees, that was his business.

So, we have a much larger willingness to accept, but

the interesting thing is that the hypothetical willingness to pay to take the tree home was two point one times the actual willingness to pay among the same subjects in a real money auction.

So, the hypotheticals were overestimating by a factor of 2.1 the actual willingness to pay experiment.

Now, one thing bothered us, and that was as the subjects read about what was going to happen, we can clearly see that they were getting agitated. Here is this cute little tree.

We are talking about killing it. So, we noticed that people got upset. In one case, one of the subjects in the experiment, fortunately after people had submitted their bids, otherwise the experiment would have been scrapped, after people had submitted their bids, this young man raised his hand and asked, "Is it okay? Could I buy all of the trees?"

We took the guy outside of the room, and we explained it, you know. The U. S. government really needed to know if

people cared to save the lives of trees, and it was the Forest Service that was behind this atrocity, and he said, "I think that is a really valuable thing to be doing," and he went back in the room and kept his mouth shut.

(Laughter.)

People were actually agitated. What we decided to do since people were agitated, would be to go find other staff members of the University of Colorado, okay, and ask them their values, their willingness to pay if we were going to kill the tree.

Sitting at their desk, where they weren't worried that they were really going to have to do this, and their values were ninety percent of the market values we got of the seventy-two actually that were obtained in the lab.

So, I will let you make what you will of that. I think that is clearly interesting that people apparently were quite agitated at the prospect of doing it.

We can use insurance policies in a lab experiment, and we get one point two, okay? We can use different

probabilities in insurance policies. This in the Journal of Behavioral Decisionmaking, in the Journal of Risk of Uncertainty. We get varying rations of 2.5, 1.8 and .8 as we go from low, median, and the higher probabilities.

There is one infamous art painting. We sell an art object, and we use a hypothetical, and he gets 3.9. One of 16th Century math, he gets a ratio of 9.1. There has to be an outlier in the '87 experiments, right?

The median of all of these studies is 1.65, and that is open-ended willingness to pay, compared to the real auction bids in an essential palatable function.

So, much for the notion that open-ended willingness to pay underestimates values. Have we heard that today? We have heard that twice. Okay.

Now, studies that compare open-ended willingness to pay with dichotomous choice for something that works like the referendum. Solice, Stolen and Chevez, studied recreational boating and used both a referendum CV and an open-ended CV.

For the three lakes, the ratios are between the two

methods 5.9, 4.8 and 9.5. Walsh, Johnson and Kean did a study of many CV studies, comparing many methods and in a regression analysis they show that open-ended willingness to pay is exceeded by referendum approach by a factor of 1.3.

A very nice study, Keeley and Turner, show that for a public good, the ratio is 1.4 to 2.4, and for private good, 1.0. The median of these is 2.4.

So, referendum exceeds open-ended willingness to pay, the median estimate is by a factor of 2.4. We have a factor of 1.65 for open-ended willingness to pay on real values in real auctions. If I multiply these two numbers together, I get a bias factor of 4.0.

Experimental economists tend to be somewhat suspicious. Okay. This is a lot of work, but what is my conclusion? Howard Schuman paid too much for his dog. It is a lot of work to show that.

Experimental economics has a lot to say about posted off the markets, okay? And I have a lengthy viewgraph that has a nice quote out of Fulton Davis' book, and basically it

says posted off the markets in round one, like a CV survey, are very inefficient. About fifty percent efficient, and people always over bid.

In other words, posted off the market in practice are not incentive compatible unless you give people lots of grounds of experience.

Now, I think this would be a surprise to the NOAA panel, and I think it is probably a surprise to my colleagues in contingent valuation, but experimental economics suggests ask people for a willingness to pay; don't ask them the dichotomous choice question. It is a very clear conclusion that comes out of a lot of cited articles, okay?

Now, very quickly I want to address the issue of mail versus in person. I am trying to hurry too much.

The real question is how much should we spend on CV studies? And that question is if in-person interviews cost two hundred dollars, and I have heard rumors of five hundred dollars, okay? And mail surveys, or mail telephone surveys

cost significantly less than fifty dollars, that is a very large cost difference.

So, the real question is: Is the increased accuracy, if there is increased accuracy within a certain area, is it worth the cost? That is really the question we have to address.

The sample sizes that we are going to have to employ in a CV study are likely to follow the NOAA regulations. If we use willingness to pay questions, which I just argued may well be actually more appropriate than referendum questions, we might get away with sample sizes of two thousand. If we have to use stoke tests, where we have to move up in the size of the commodity and down in the size of the commodity, that is when you reach three different surveys and sample sizes are probably about seven hundred a survey, you show significant differences.

If we do this in a referendum, where we may need five different values, we might need sample sizes of ten thousand people in a survey. I am happy to argue about that.

But that means the cost for a willingness to pay mail survey might be a hundred thousand dollars, a mail version of a referendum survey might be five hundred thousand, and in person it would cost four hundred thousand, or two million if we use a referendum approach. And that is just for data collection, without pre-testing or any other design effort.

So, these costs are simply outrageous, and the question is: Are they really worthwhile? Are there things that we can do to solve the problems that Howard Schuman has raised concerning mail surveys.

And I got this one right side up. Now, I recommend you do look at Don Dillman's letter, because I think he does address a number of Howard's concerns, and I should point out that Don Dillman is chief scientist of the U. S. Census, so at least he has a place to stand, and I don't want to engage in a debate about Howard, since he is the survey expert, but I think Dillman is in a good position to do that.

One very serious question that Schuman raises is that there is no adequate listing for national and international populations. That is a serious, a really serious problem. And another serious problem is non-response. However, there are solutions in the literature to each of these problems. For example, Lunus and Harman in a study in California, used random digit dialing to reach a random sample of all people.

The problem with telephone listings is that many people don't list their phone numbers. They have unlisted numbers. Random digit dialing reaches everybody, and then they simply got people to agree to fill out a survey, and they got eighty percent of those people to agree to receive the survey in the mail. So, they got the addresses over the telephone.

In a recent study in Montana which has been mentioned, we offered respondents twenty dollars if they filled out a survey and returned it to us. We got a very high response rate. We had a very lengthy and complex survey.

So, if you put these two together, the notion of using random digit dialing to get a random sample, with substantial responding conversation, I would argue that you can overcome many of Howard's legitimate objections.

So, the issue here is that with costs of up to five hundred dollars a completed in person interview, there are a lot of things that you can do to find a good sample.

There are lots of clever things that you can try out.

So, my conclusion then is, first, as I think it was made obvious, by the question of open-ended willingness to pay versus the referendum, that there really aren't any uncertainties which remain in determining the ideal CVM. And I would hope that regulations such as those proposed by NOAA and DOE recognizes those uncertainties, and don't limit the methodologies that we can apply as much as the initial drafts of the regulations in effect did.

But I also want to point out that enormous uncertainties exist in all areas of public policy research. If you are trying to get a toxicologist to look at the

mortality of trout in response to exposure to heavy metals, there is an order of magnitude of uncertainty. So, order of magnitude of uncertainties are normal in litigation, and they are normal in the public policy analysis.

So, simultaneously while saying I think the uncertainties are large, I think that does not mean we throw out what I would call order of magnitude estimates of images.

MR. PORTNEY: Bill, thank you. It may well be the case that Bill can be heard by four hundred undergraduates without a microphone, but they are much younger than we are and haven't started to suffer hearing loss yet.

Our next respondent will be Bill Desvousges. Bill is currently Director of the Research Triangle Institute's Natural Resource Damage Assessment Team. An economist by training, Bill has been working on natural resource damage issues for I guess at least the last five years in some particularly celebrated cases, and I am particularly happy to have him take this opportunity to respond to Howard's

paper.

Bill?

DR. DESVOUSGES: As I have looked at the schedule here, I notice that I am now down to three minutes left.

In order to try to save time here, to keep us on track, I can probably help do that by not responding at all to the third study.

I think that when we did that study X number of years ago, none of us had any idea what commotion might result from it.

I do think during the course of today we heard some very interesting and challenging remarks that have been made.

I think that Professor Schuman's comments are particularly relevant to many of the issues that I think we need to discuss.

One of the things I believe is extremely important in any debate is to clarify the terminology that we are using in that debate. I noticed a number of different assertions

that were made back and forth this morning that dealt with a number of tests or whatever that have been done on various things.

Whether we have no tests, or a bunch of tests or anything else. In order to answer that question, we have to be very clear of what the data base is in which we are trying to draw conclusions.

Are we talking about use values, are we talking about studies that are some combination of use studies, or are we talking about some studies that are largely nonuse studies.

I think depending on the answer to that question one can say very different things about what we know about the performance of contingent valuation in the number of studies that have been done.

So, I think as you compare, and hear people make comparisons, it is very important to keep in mind what the data set is on which they are trying to make those comparisons.

If it is use factors, I think many of the issues, many

of the concerns that I had about contingent valuation then, or at least they are certainly minimized, compared to situation where we are talking about nonuse values.

I think that we really need to make that distinction from the outset, and that we also need when we are trying to draw inferences about nonuse studies, to make it clear when we are using data that are based on studies about largely nonuse values versus ones that include larger use values.

The second point that I would like to also add is that we have heard a number of discussions, very interesting ones, about golden retrievers. I happen to own one. I made a slightly different choice when I went into the market, and I think this also helps to show you the thing about markets versus CV studies.

We thought we were going to buy a putty as well. However, when we got to the breeder's house, she said, "I also have a one year old here, that has come back to me. We also have puppies here, but you might want to take a look at this guy, too."

And we took him for a walk and we thought about it for a while and we said, "You know, it is really not so bad. It is a hundred bucks less for a year old dog, and we don't have to housebreak him."

So, the bottom line here really is that the market place really embodies varying attributes that are out there.

In a CV study the only attributes that are available to the respondent are those that are provided by the person who makes up the survey questionnaire. The second distinction that I would like to make for you in terms of comparing markets and how people behave in markets is a very important one as well.

And that is, markets punish when you make mistakes. CV questionnaires don't. We have heard a lot of different things about rationality, irrationality and everything else, but I think it is very interesting, and there is a lot of appeal to it. Gee whiz, people do silly things in the market. They really do. There is no question about that.

But, if you are in a market place, if you make a

mistake, you are going to pay more than what you would otherwise pay. We don't have necessarily that same kind of correspondence when you are talking about CV.

All, right. So much about polemics there at the outset. Let's see if there are a couple of other things I can add to it.

One, is I think it is very hard for us to make definitive conclusions about performance one way or another based on literature that we really do have that deal largely with nonuse commodities.

In my paper, I have identified thirty-seven studies that are published or closed to being published, such as the Alaska study. That is the thirty-seven studies that I think are there.

Now, obviously, I may have missed some. But if we are talking about those thirty-seven studies, it is interesting to look and see whether or not one of the statements that Professor Schuman made, the literature is very rich for making tests of hypotheses.

So, I think of the thirty-seven that tested for differences related to theoretical validity. In addition, when you are talking about studies that have done tests, there is another important characteristic to look at, and this also relates to an earlier comment this morning.

Are the tests, or were the tests done with split samples. It is a very different test when you use a split sample than when you ask the same person the same kinds of questions. A lot of the scope tests that have been done have been done on the same individual.

I think one of the things that we try to do, and actually I guess it has been the thread in many of the CV studies that I have been involved with over the thirteen odd years that I have been doing it, is to try to use experimental designs to inform and motivate the discussion.

It is very important when you are doing experimental design to pick things that might make a difference.

For example, Professor Schuman says, "You picked a bid that was unreasonable, implausible." I agree a hundred

percent. That is why we picked it. We wanted a bid that we felt would choke off demand to virtually zero.

So, we took a bid that was twice what we had in our retest, and said, "This ought to do it."

Obviously, it didn't. Thirty-three percent of the people said yes to that bid.

Now, does that make our study implausible that someone else who doesn't pick a large bid, their study then becomes plausible? I think it is really an open question. When you are designing experiments, what is the ranges of values that you are going to use in your experimental design?

When we first tested back in 1982 in the Monongahela Study for starting point bias, one of the open questions was does it exist? Well, most studies had taken bids that were fairly close to each other and said, "No problem." We took starting bids five and -- I can't remember now. Twenty-five or five, somewhere in there -- and a hundred twenty-five dollars, a long way apart.

We found that it is a starting point bias.

So, once again I think it is very important when you try to look at these issues that you use an experimental design. That experimental design must involve treatments that are going to say something about the issues that you want to try to evaluate.

So, I have a little thought here to keep in mind as you are trying to work your way through what you heard today.

One of the other things that I think we have heard a lot about is open-ended questions, and whether or not they are reliable or non-reliable.

I don't particularly have a big stake in this issue, but I do think this is one of the places where do have at least some evidence on that about reliability, using the narrow term in terms of two people answer consistently. John Loomis did a study of Mono Lake where he did retest for reliability. The type of question format that he used was open-ended. He found a correlation of roughly about seventy percent in his responses. The study was done at least nine months after -- the second follow-up was done at least nine

months later.

Kevin Boyle did a study of moose hunting where he had a very unique experimental design, found essentially the same kind of situation.

So, I think the evidence is a lot more mixed than what we heard today about what you get with open-ended questions.

The next point that I would like to make is what I call the rush to referenda. If I think that we have been on a train for the last several years, without taking a good, hard look around us, it is the referenda train.

It scares me. I think that it is extremely important when we start adopting question formats of any type, and I won't mention what we have done to favorite vehicles lately, because we have virtually ignored that topic, but when we start to pick referenda formats without thoroughly evaluating the properties of those formats, I think we ought to be concerned, whether it is public policy damage assessment or whatever.

I happen to feel like, contrary to Professor Schuman,

that CV studies are not referenda votes. Let's be realistic. In a referendum, voters pay. Do CV respondents pay? I think not.

Secondly, the dollar costs that are provided in CV referenda are the tools of the survey designer, not the actual cost of the policy program.

Third, when we are doing a referenda, frequently the main objective is to determine whether or not fifty percent or more of the people will support it. That is the goal.

In a damage assessment in particular, we are extremely concerned about the absolute magnitude of the dollars that are on the table. I think that is a different standard than for other applications as well.

Finally, I think Professor Schuman really started to make some good points about this point, but I want to underscore it, and that is the access to information in a CV referenda is completely controlled by the person who designs the study.

Is that the case when you go to the polls to make your

vote? Not at all. You have the choice, if it is important to you to decide to acquire more information. You have the choice of not to vote at all, as he points out. You also have the choice to get as much information as you possibly could have in order to make an informed decision.

Or you could do like I do and read the News and Observer and see what the editors there say you should do, and do just the opposite.

(Laughter.)

So, there are different strategies that voters use in terms of trying to do this. And I think as we start down this rush to referenda that we really ought to take pause and ask ourselves what is the performance of the question format? What are the true implications of what we are doing? Which features are important, and which are not.

The last thing I would like to do is to leave you with a few thoughts for future research. I will try to keep to the theme of the conference for a change.

The first is that I think it is extremely important

that we focus more attention on comparisons of hypothetical natural studies. However, I have a proviso on those kinds of studies that I think need to be done. If they don't include some component of non-use values, their value is going to be limited, because I think the conclusions that we can draw about performance related to market goods, compared to goods that are largely composed of peoples non-use values, are very different, and that I don't think, contrary to some of my colleagues, that you can have one bias function that is going to explain all of those.

I think it is a desirable goal to be looking for those kind of things, but let's keep in mind that the type of commodity could be very important in terms of how people respond.

I happen to think it is very much an open question as to how survey mode matters, and whether it matters.

I think that there are a number of issues that are wrapped up in that. Whether you are talking about response rates, whether you are talking about the valid sample

frames, those are important features as well.

In addition, what you want to know is does it matter whether you use the personal interview or not. How many times in contingent valuation have we tested that question? I don't know the answer to that, but certainly I have had a hard time trying to find studies that have done it.

Obviously, people will help me out with that. Finally, what I think we ought to do is continue evaluating this issue of question format. As you have been able to tell so far, I am not a fan of referendas.

It is kind of funny. I have used referenda formats in about five different settings, and I have found, and I am going to use the technique of personal conjecture here, I have found that when we have done things that have involved primarily a commodity people have experience with, or that have use value, we don't run into the fat tail problem that we did with the oil spill experiment.

People have a much better handle on what is an implausible value or not if they have had some experience

with it. And they go, "Whew, no way I am going to pay five hundred dollars for a radon test kit if this should come into my house. No way I am going to pay more than twenty-five dollars for a test kit for radon.

There is a market out there, there is experience with that.

The other component that I would like to add to this question -- I think as you hear Jordan Louviere's paper, I think he is really raising some very interesting questions here about conjoint.

I think conjoint -- once again, it is not a panacea. I don't want to see us have another train wreck here on another question format, but I do think that there are a couple of features of conjoint that may help us to start focus on what I think is an important issue.

First is the conjoint requires people to make trade offs. We have heard a lot of arguments about -- it is positively much easier to answer a yes/no question. Do we want cognitive ease to be the primary determinant, or do we

really want serious research of preferences to be our main concern?

It has also been a little worrisome to me that when we do a CV study we spend 45 minutes telling people about the commodity, and then we give them 38 seconds to give an answer, yes or no, to this one number, and that is it.

I think it is very important for us to begin to do some more of the things like Dale Wittingham and colleagues did, looking at issues about time to think, about information, about when you provide the commodity, when you ask for the value, and what question format and survey mode that you might use.

Thank you.

DR. PORTNEY: Bill, thank you very much, and I especially appreciate your efforts to keep us on time. I think we are back on track here, and I am grateful to you.

Our next respondent is Jonathan Baron. Jonathan is a Professor of Psychology at the University of Pennsylvania. His research interests have explored individual rationality,

the use of futuristics in decision-making, and recently individual attitudes toward risk.

I am pleased to turn the floor over to him.

DR. BARON: I am going to do my part to help with the time problem by omitting the many points of agreement that I have with Professor Schuman, and get right to the disagreements.

The main disagreement I have is about where the problem may be. Not the problem. I am not denying that better survey methods will help. I am suggesting that there are other possible more serious problems with the validity of contingent valuation.

I want to talk about three of them. One is a well known problem of insensitivity to scope, the second is the idea of fairness or fair prices, and the third is the problem of fundamental value.

I think when research is done on various manipulations of contingent valuations, such as open versus closed questions, I think it should not only try to see whether

those effect the prediction of market behavior, but also see what effects these manipulations have on these problems, in particular it might be nice, for example, to ask whether open versus closed questions affects insensitivity to scope.

There is one attempt to do that, but I don't think it is very good.

So, let me get to these problems. In insensitivity to scope problem is not just the embedding factor, but there is also a couple of other forms. There is the adding up effect, talked about by Diamond in which -- in his politics, in which the willingness to pay for A, and the willingness to pay for B, given A, don't add up to the willingness to pay for A and B together, or anything close. I have done studies like this where I get fifty dollars, fifty dollars, fifty dollars. Fifty dollars for A, fifty dollars for B, and fifty dollars for both.

And the third effect is what might be called the quantity affect, which has been probably known about for a long time but just recently paid attention to, particularly

by Michael Jones Lee. The quantity effect is my term.

What Jones Lee does is ask people their willingness to pay, for example, for safety devices that reduce risk in say automobile accidents by varying amounts, and Jones Lee takes great pains to convey to people the amount of risk reduction. He has, for example, a card with ten thousand dots on it, and indicates the risk by shading in dots at random.

And this is all in personal interviews. And he finds that many people, even within subject design, when people are asked about different amounts of risk, have the same willingness to pay for different amounts of risk reduction, and that happens both within and between.

Even when you look at average figures, or if you eliminate those people, you still get under-sensitivity to risk. That is, the rate of substitution between risk and dollars, risk reduction and dollars, isn't constant as a function of the amount of risk reduction asked about.

People are willing to pay whichever way it is, but

basically they don't adjust their willingness to pay enough to compensate for changes in risk. I don't want to try to say that much, I get it backwards.

(Laughter.)

The problem with this kind of insensitivity to scope is that in the end, I think, most CV methods -- and I may be wrong about this, but I would like to hear it -- most CV measures really want something like amount of dollars per unit, of some sort of unit.

For example, Jones Lee, his original idea behind this research was to give the British Ministry of Transport a figure on how much money they should spend to save a human life on the highways, and he wanted them to rationalize all their policies based on that figure.

So, what he wants is dollars per life, and he wants to infer that from the willingness to pay for risk reduction.

Another example that may not seem like it is dollars per unit is in tort cases, or cases involving damage penalties like the Exxon Valdez oil spill. According to the

economic theory of tort law, which may be wrong, but the way this is supposed to work is that when a company or an injurer is penalized for causing some damage, they are supposed to say to themselves, "Oh, that is the cost of the damage."

Now, if I am going to prevent something similar from happening again, if I can spend less money than this costs, then I ought to be able to do it.

So, the policy of Exxon, or whatever company over the long term, would be set by figuring -- by using the damage penalty as an estimate of their future expenses from lawsuits for oil spills or whatever.

So, again, it comes out to dollars per oil spill, so far as Exxon is concerned.

I am sure that is a rough approximation of the truth, but it does suggest that what people really want is dollars per oil spill, not a one time, one shot thing.

So, if people are willing to pay the same amount to prevent one oil spill or to prevent ten, what is the true

dollars per oil spill? And also, that leaves us the question of whether people are really rational in another sense that I want to talk about.

But before I get to that, I want to talk about what is going on with self sensitivity, which I do think is a topic worthy of further research. I am doing some right now, but I am simply not going to get all the answers in the next year.

There are several explanations that have been proposed. I will just go through them quickly just to give a flavor for how many different ones there are.

One possibility proposed by Kahneman and Knetsch is that it has something to do with moral satisfaction. I don't think they said this, but I think there are two versions of that hypothesis, the moral satisfaction account.

One, is that people think of willingness to pay as something more like a charitable contribution, in which case the amount of good that gets done is proportional to the amount I contribute. It doesn't matter whether I think of

that contribution as going for, say, one village in Africa, or spread randomly over all of Africa, with all other -- pulled in with all other contributions.

It is still going to do the same amount of good.

So, the size of the good wouldn't matter in that case, so you might think of that as a kind of major scenario to which people would refer this contingent valuation question.

A second version moral satisfaction is the warm glow account, which is that people get utility or satisfaction or something from the activity of contributing to a cause.

It is not that the money is contributed. It wouldn't make them as happy if somebody else gave the same amount of money.

If these account for the embedding effect, you would think that that effect might go away when you use other kinds of methods. I don't want to say these are definitive results, but I find equal embedding effects when I use just simply ratings of how good it would be -- how good it would be to take care of this problem.

Again, in just ratings of goodness, or seriousness, or many different kinds of ratings, I find equal embedding effects, equal in sense to scope, than what I find with willingness to pay.

I also find equal insensitivity to scope in most studies when people are given a trigger price, so that they can't think of their contribution as analogous to a charitable contribution. At least they shouldn't think of it if they pay attention to what I say. I am not sure that they do.

So, if I tell them for example that here is a given good, a ten percent increase in something, and it will either occur or not, as a function of whether fifty percent of the people or more are willing to pay what it costs per person, that ought to get people away from thinking of the charitable contribution idea. It doesn't have any effect on the whole, but it may be that they are really not paying attention to what I say either.

There are three other explanations of embedding, or

scope insensitivity. Actually, four. Budget constraints.

People may think fifty dollars is what I have to spend right now, so fifty dollars for everything. This is all very important, but that is my budget.

That predicts that you wouldn't find scope insensitivity in willingness to accept questions, as opposed to willingness to pay. I have found scope insensitivity in willingness to accept questions, but in only one study. So, that casts doubt on the budget constraint story.

The second additional explanation is some kind of general mental accounting story, that is both for pain and accepting money, I have a kind of limited amount of concern.

It is hard for me to put this into words, but there might be some more general kind of mental account that people have for different kind of things. That applies both to willingness to pay and willingness to accept, such that they operate according to a principle of declining marginal utility.

Okay. A third is something like availability. Mitchell and Carson proposed something like this. Schulze proposed it in his paper. Namely, that the good that people are asked about somehow stands for the whole set of goods, either because people think you can't provide the little good without providing a much bigger good, although I don't think that is not plausible in all the cases. Or in some other way, in which the little good stands for a larger good.

A final explanation, which I think is kind of interesting, because I haven't seen it before, is that people make their willingness to pay responses on the basis of their view of the importance of the problem rather than the amount of good.

I had one subject in a study where she gave equal willingness to pay figures for several different kinds -- she gave willingness to pay figures for several different kinds of goods in one study. Each good was presented in a large magnitude form and a small magnitude form.

She gave identical willingness to pay responses for both the large magnitude and small magnitude versions of the same good. But she varied from good to good, so she said five dollars for a little of good X, five dollars for a lot of good X, ten dollars for a little of good Y, ten dollars for a lot of good Y.

And then I said, "Why did you do that?" and she said, "Oh, well you know, if I raise the five dollars for the large amount of good X, I probably would have had to raise it above the ten dollars that I paid for a small amount of good Y, and I don't want to do that because good Y is more important than X."

And it could be that people are thinking about this even when the other goods are not presented. So, there is some kind of overall importance rate.

Okay. The sensitivity of fairness is the second problem. Three, I guess. I will just give one quick example. You ask me how much I am willing to pay for a radon test kit. What I think about probably is what does a

radon test kit look like? Oh, it is a little canister, it has some chemicals in it.

It really couldn't be worth more than twenty-five dollars. What I don't think about is what it would be like dying of lung cancer. What is the probability that I would die of lung cancer with and without getting the radon kit. And so on, and so on.

So, essentially I am asking -- what I am answering about is what is a reasonable amount of money to spend on this good, rather than how much utility do I get out of that?

Okay. The rest of what I wanted to say I will just -- I had a lot to say. One point -- reduce it all to one point.

There are two general issues invalidating contingent valuation that I think are somewhat confused. One of them is -- well, maybe not confused. One is the question of whether contingent valuation responses predicts real world behavior, like market behavior.

The second issue is whether contingent evaluation does justice to our fundamental values, our goals, our utilities, in some sense.

Now, the answers to these questions are not necessarily the same. It could be that contingent evaluation does predict market behavior perfectly, but both are out of line with our fundamental goals as Professor Schuman suggested was possible.

It could be that it doesn't predict market behavior, but it is in line with their fundamental goals.

In my paper, I suggest several ways of thinking about the question of whether contingent valuation does satisfy our fundamental values and goals in the sense maximizing utility.

It seems to me that -- at least for some purposes that is the question we want to answer because of this. If all it does is predict market behavior, a market-like behavior, but it doesn't achieve our fundamental goals, then it becomes a kind of black box, namely we accept market

behavior as a criterion even though it is irrational. We just respect it because it is what people do.

So, contingent valuation acquires value because it, too, it simulates that black box. We put in information, and we get out a price.

I want to suggest if that is all contingent valuation is, without at the same time achieving something more fundamental, it is not going to stand the test of time because there are other black boxes that we already have.

We have things like juries, for example. We put in a lot of information, we get out a damage award. We have legislatures, we have government regulatory agencies that do these things.

We don't look too closely to see how any of these things work, but if contingent valuation can't acquire some justification other than being just another black box, I think it is going to be a problem.

Okay. I will stop.

DR. PORTNEY: I want to give Howard Schuman an

opportunity to respond briefly here.

DR. SCHUMAN: I don't really want to respond. I just want to reiterate one thing. There is a big difference between doing experiments in order to test particular points and there one can often not use a full probability sample from the general population and so forth.

And the other is to come up with a dollar figure, as was attempted in the Exxon Valdez case, and there I think you do need to go through that. And it is going to be very expensive.

Every month the census bureau gathers fifty thousand or more cases, I guess, to get an unemployment rate. That costs a lot of money.

If you are trying to do something that requires a lot of precision and needs to be done well, and is going to have an important role in litigation and so forth, then it is going to be very expensive.

For other things, I and other people often take short cuts, and that is perfectly legitimate.

DR. PORTNEY: Thank you. The temperature in this room puts me in mind of the story of the two guys who were having a tough time in the aeronautical engineering class. They failed all of the exams. The final weight of the grade rests solely on their final project, so they think very hard about this, and probably go to their professor and say, "Look, we know we have had trouble in this class, but we have been working hard on our final project. We are going to attempt the first man landing on the sun."

Well, the professor shakes his head and says, "You morons, it is very hot on the surface of the sun, if you haven't noticed. You will be immediately incinerated."

They said, "Yeah, we knew you were going to say that, but we are going to land at night."

Maybe if we came back here at night this room would have cooled off sufficiently. The best we can do is take a brief fifteen minute break.

Be back here at ten after three to begin the final session of the say.

Excuse me -- come back for open discussion. I am sorry.

(Whereupon, a recess was taken at 2:55 p.m., to reconvene at 3:15 p.m., this same day.)

DR. PORTNEY: I have two brief announcements I want to make.

The first pertains to those of you who have special meal requests. This doesn't mean you can ask for a lobster to be jetted in from Maine, but if you have questions about what we are supposed to have for dinner, and do have legitimate special meal requests, if you will just check with the desk right outside, between the two doors here, especially check with the woman in the purple dress and let her know what the problem is we will try to make arrangements to accommodate you.

Second, if you look at the program for tomorrow, there is a misprint there. It says that Wally Oates has been allotted an hour and fifteen minutes for his paper. Even by the generous standards of this conference that is a lot, so

we are going to trim Wally back to just forty-five minutes, and we are going to do that by starting at 9:00 tomorrow.

There is a lunch that will still end this thing, but we will have the luxury of a bit later start tomorrow.

So, tomorrow morning's program will start at 9:00 rather than 8:30. Please keep that in mind.

I want now to give you opportunities to ask questions or make comments directed at Howard Schuman, Jonathan, Bill Desvousges, or Bill Schultze, if he is still here, so we will start with Robert Mitchell.

ATTENDEE: I am Robert Mitchell. I would like to make one comment address to Howard Schuman.

REPORTER: Sir, I cannot hear you.

DR. PORTNEY: Yes, it is hard to hear. I don't know if the mike is dead.

ATTENDEE: My name is Robert Mitchell. I have a comment directed at Howard Schuman. The comment to Howard is in reference to your last statement, where you offer the horrendous fear that everyone -- perhaps thirty percent of

the people will say yes to any amount.

There have been a number of studies which have used the dichotomy of choice format, and in which amounts varying from five dollars to seven hundred dollars have had less than ten percent of the people saying, "yes," to them.

Let me just mention two of them, since they involve different values. One is a study by Falcon Stall, on the Whooping Cranes, at a hundred and thirty-five dollar price, less than ten percent said, "Yes."

Whitehead did a study on preserving Loggerhead Turtles, building a habitat in North Carolina, and a hundred dollar price, seven percent said, "yes."

As it turns out, we went back to the Alaska data to our pilot studies, and found that in the very first pilot study we did offer people a thousand dollars as one of the prices. In that case, we had zero percent who said, "yes."

My particular desire for future research, I think this year's number you have raised, which relate to the issue of non-entities, and to the issue of interviewer effects and

this, of course, is something that the panel emphasized, I think these are two very, very important issues that deserves serious treatment, looking at such things as different ballots, and looking at such things --

DR. SCHUMAN: I am not sure there is really a question in there. It is interesting about the studies that didn't produce substantial percentages for what seemed like incredibly large amounts. Dr. Desvousges might want to address that.

I talked about that, because on the one hand his study struck me as showing such a high percentage of agreement to a thousand dollars a year. At the same time, I said it is not a sample -- it is not a random sample of any known population, and the conditions were very different so that it is impossible to draw conclusions.

DR. DESVOUSGES: The first question -- I think this is really related to a general issue in the referendum, which is the choice of the bid structure, and if you find bids being chucked off at a thousand, but yet with the highest

bid that you used in the final survey of only being a hundred to a hundred and fifty, and still have a full third of the people saying, "yes," there, perhaps that suggests that the bid structure issue in terms of what you pick and how you pick it is a really important one that we are still grappling around with in terms of trying to figure out what matters and what doesn't.

I also like to point out that we are not the only people that found fat tails, and some concerns about whether people are anchoring on some of these questions, or whether surviving.

Barbara Cannon's paper, I think, has raised some very interesting issues on what is going on here and how people answer these questions.

There are, I think, a number of questions that are on the table. I am not suggesting that our results are representative of any population other than the people who took the survey, because the only thing we tried to do was compare them with people -- similar people who took an open-

ended survey.

I think for our purpose that was the only goal.

DR. PORTNEY: Howard?

DR. SCHUMAN: I just want to agree with Bill, that it is very unfortunate that the Alaska study did not include a much larger value for whatever reason, even if it wasn't needed for estimation. As the other Bill said, it invested an enormous amount of money. Most of the critical studies are tabled to do that, and it could have addressed this problem very directly by including a very high value in reporting the results.

Not just a pretest.

ATTENDEE: Let me respond to that. There is a confusion that has been caused, because you are just looking at the first response. The Alaska study has double value response, which the highest value was two hundred and fifty dollars, of which the estimate at two hundred and fifty dollars is eight percent.

So, one may criticize the double value approach, but in

terms of choking off demand, if we got down under ten percent --

DR. SCHUMAN: We don't have any idea -- there are real artifacts of asking people something, and then asking them larger or smaller amounts.

So, I don't take those very seriously.

ATTENDEE: In terms of the design itself, two hundred and fifty --

DR. DESVOUSGES: May I just respond to the other point?

DR. PORTNEY: You may certainly briefly respond.

DR. DESVOUSGES: I will do it very briefly. The question of kind of split ballots and anonymities is, I think an interesting one, but I would question how you would implement it in terms of -- Howard said that, well, we had people mail back the responses, and there would be no identification associated with it.

It would be very difficult then do a comparison in terms of socio-economic characteristics if you don't know who the people are that are responding.

So, maybe you can do some compromise in there, but at least you put a number on the ballot so that you could do that, even though there might not be a name on it.

If your goal is to do some kind of comparative evaluation, you really need to know who the people are that have responded to the questions.

DR. SCHUMAN: This would just be a test of whether there are strong interviewer effects, and you would not try to identify the people.

Moreover, if someone didn't send back a ballot, I would treat it as a disagreement. If they don't value it enough to send back a ballot, that should be a no response.

DR. DESVOUSGES: One of the things I took from your paper was a concern about only having uni-varied tests, and if the only thing you need to know is the mean value that is going to come from that, you are going to be hard pressed to take into account the socio-economic characteristics, and how they might influence it.

DR. PORTNEY: Glen Harrison and then Dale.

ATTENDEE: Since this conference is about setting research agendas, I would like to talk about an issue that was raised here, and Bill is the expert in that area, and that is a research concern of mine, which is my wife's fat tail.

Namely, she has been running these, and she has a fat tail problem which is that she can run an experiment for private goods. She looked at an open-ended victory auction. She looked at an closed choice, hypothetical, and she got the classic fat tail problem. Then she did both of them real. Wham, they disappeared.

Arguably, that is just a conjecture. The fat tail problem is an artifact of staring at hypothetical behavioral interests, and that is a question of priority.

Shouldn't we be worrying about queuing effects, sampling effects, splitting design effects, population sampling, median sampling, or should the priority be let's check -- let's do a reality check to find out if it was just a differential hypothetical bias or what?

The fuller question is the relevance of doing experiments on just used goods, which I have done, and I share your remarks that those are of limited value, but let's make a point that is only an issue of research budget.

There is nothing in that design -- that is an important one and I will take the criticism, but there is nothing in the experimental methods that couldn't be applied out in the field. I think that is an issue I also was hearing from Schultze, because he has been an eminent experimentalist, but also doing a lot of stuff in the field.

What are the problems we might get in doing that?

DR. PORTNEY: Bill, why don't you take it first?

DR. SCHULTZE: There are a number of articles within experimental economics that look at incentive effects, and really I think we owe some of our interest in this to a paper that Glen Harrison wrote, where he examined the pay out structure, and noticed as you deviate from an optimal bid in an auction, that in many of the studies the penalties were very small, could be twenty to thirty percent off of an

optimal bid, and you only were losing a few cents around.

So, that leads us to the following question: In the real world, he draws an analogy to the shape of a behind. Are pay off functions fairly flat? And this goes back to psychologist named Edwards, who sent graduate students out in the late '60s to see what the shape of pay off functions were.

It turned out that pay off functions in the real world are pretty flat.

In other words, competitive firms can make big mistakes, and they don't pay huge penalties, okay. So that is kind of interesting.

This might mean that people can be pretty sloppy in their bids and not have large losses in experimental auctions.

So, there are a number of experiments, then, that try to change the incentives in the sense of how much does a mistake cost you?

What we find, of course, is that as the incentives

increase for making a mistake, the distribution of bids narrows and becomes more peaked. Is anybody following me so far?

So, we tighten up the distribution of bids as these percentages become greater.

One of the monetary incentives in the CV, where the incentives have gone to zero for accurate bidding, except when we run a totally hypothetical laboratory experiment, where we take incentives to zero, and we can do this starting with a ten dollar, five dollar, and run it all the way down to zero, and the behavior does not degenerate to noise.

We don't get a level function as a distribution. We just get, again, another increment increase in the variance in the bidding.

Unfortunately, the shape of these distributions all show thick, right hand tails. So, we have the increase in variance as we lower the incentives down to zero, and we do not have a normal distribution of error. We have a big,

right hand tail, whether there is real money involved or not.

In fact, starting in 1990, we started box coding our CV bids, arguing that there was a skewed distribution of hypothetical measurement error, and we found the coefficients were almost uniformly not significantly different from zero, which means it appears that bidding errors are logged normally distributive.

That means an increase in variance will increase the mean, okay?

So there are statistical procedures that can be used in contingent valuation, namely take the geometric mean is the advice that I would give CV researchers, and you may be correcting for this hypothetical bias, okay?

So, that is a suggestion, and perhaps I should stop there.

DR. PORTNEY: Dale Whittington?

ATTENDEE: University of North Carolina. I have a question for Professor Schuman, but first just an

observation or an initial comment.

A series of studies, we never had this fat tail problem. Bill is right, it was use data. At the same time it gave people time to think, even the small tails we had disappeared.

So, it is worth considering other kinds of treatments that might improve the quality of CV. My question to Professor Schuman is Michael Hanemann asked this morning, or he said this morning that if we just treat no responses appropriately -- he didn't tell them how to treat them appropriately.

I would like to ask you what you think the agenda is, really, for the treatment of no responses? And how can they be treated appropriately?

DR. SCHUMAN: It is a moderately important issue, I think. I mean, it is true that in a national survey on anything difficult a fair number of people will say, "don't know," if it is legitimate. And it also has to be legitimated for the interviewer, because interviewers are

usually trained to push people to give a response.

So, it serves a double purpose to build it in, and I think it is a useful thing to do.

I am not sure if I have addressed your question or not.

ATTENDEE: I guess my question is we know that these, "no responses," are not randomly distributed --

REPORTER: Excuse me, sir. I cannot hear you, sir.

ATTENDEE: The, "no responses," are not randomly distributed in the population, so I am just interested in what has been done in survey research literature to deal with this problem?

DR. SCHUMAN: In a case like this, it is actually the least educated who are most likely to answer the question even if they don't know, really.

So, yes, it is unlikely to be any kind of a random sub-sample. And I would drop them out of the population, their proportion out of the population.

DR. PORTNEY: Mary Jo, do you have a question?

ATTENDEE: What we are concerned about this valuation

method mostly because we are hoping that the value estimates would be used for a very specific purpose.

In our case, we are hoping to use this information to make intelligent decisions about how to allocate resources for environmental protection, so the end result will be that allocations of actual dollars would be made, and in our case we are talking about billions of dollars for environmental protection activities.

And that means that somebody is going to be paying that money.

As for environmental protection activities, usually we assume -- as probably we have shown in a number of cases -- that means higher prices for goods and services, and the taxes that we all pay.

So, how is it -- Bill, let me put my question to you -- how is it that you can say that people don't have to pay for these things? We are really using the numbers to get values that will then help us make a decision.

Then shouldn't we be able to design mechanisms that

convey to people what it is going to mean to them in terms of a higher prices. So, I am not sure of the incidence of payment that runs to the national resource damage system.

In our case we are pretty sure that the price is higher taxes for school services and that sort of thing.

Why are we saying that people don't have to pay for these things.

DR. SCHUMAN: I guess there are a couple of response, Mary Jo. The first being is that I have never felt that damage assessment was that much different in some ways than public policy situations, because certainly what is at stake can be sizeable in terms of a regulation on hazardous waste producers or whatever, and the consequences of that are equally great to the specific liability of an individual firm or group of firms and damage assessment.

They are both important.

But there are some distinctions. The first is that I think for the purposes of usually choosing between a regulation, the absolute magnitude is less sensitive than

the relative magnitudes of the different options that are out there.

Whereas in a damage assessment, the absolute magnitude is the number that goes on the check.

The second distinction, and I think this really relates to your point, is that my concern is despite the reality that people may be paying for a lot of these things. Frankly, they are paying in damage assessments as well, because certainly things will be passed along.

It is whether or not people can perceive that connection to the payment vehicle, and from the payment vehicle that they are really making the payment, and I think that there are several reasons why we might be concerned about that.

First, is whether or not they would fully appreciate the amount that they are actually paying that way. That over a period of time small incremental amounts on lots of different purposes, do people really have a good sense of what is really at stake and how much they would be paying,

and whether or not they really do a calculation as to whether -- that is something that is really equal to the value that they would place on what they gained from the regulation.

So, I think it is important, but I do think there is some distancing there that people have that still leave me concerned as to a situation where they would really be making a payment at the same time.

DR. PORTNEY: Jay?

ATTENDEE: Jay, Ohio State. I have two method questions, and the first is that I have recently seen a paper using the random response method, and Professor Schuman may be familiar with this in terms of his research, which is used for very sensitive subjects like cheating or drug use, where the respondent answered the question about, say, contingent valuation, free choice, or the answer might well have been the day their mother was born, and nobody knows what answer it was with this survey. Had some value based on likelihood.

In one study I saw that used that, found that the non-random response was over-estimated by an order of a quarter magnitude with the random response technique, and I am curious if you saw that as a potential abnormality to future research.

The second part I guess is aimed at Bill Schultze, in that do you see, using lab experiments at all, to come up with the distribution that could potentially then be used in the discreet choice mechanism, rather than going to a focused group and coming up with a hypothetical distribution from questions and answers.

The focus group -- and it gets a little dodgy when you are talking about public goods, but is there a way to incorporate the actual bids from the experimental auction into the field surveys, so that you actually would have potential distributions over different trials.

DR. PORTNEY: Howard, the random response?

DR. SCHUMAN: Well, as you said, this has been mainly used with drug issues or crime, where there is a real

sanction on the person for admitting something. It would be interesting to try.

But I don't think that the interviewer effects here are of that magnitude. They are probably a more subtle kind of thing, trying to please the interviewer. So, while one would never say that research is not needed, you get paid to do it.

I doubt that that is the avenue to go. I think better is to do something where you remove the interviewer another way, by having people mail things back.

DR. PORTNEY: Bill?

DR. SCHULTZE: I hope I understand the question. We have done one experiment which is in the table that I put up, and I believe I actually cited that one in the paper. It is Erwin, McCullen and Schultze, 1992 Journal of Behavioral Decision-making.

We ran a lab experiment with a victory auction, and we did it for real for fifty and one hundred rounds in two separate treatments. There were one in one hundred odds of

a real forty dollar loss, and they were buying insurance policies for each of those rounds.

So, we had fifty trials and one hundred and fifty trials.

In the real auctions, people started off considerably above the expected value, which is typical for low probabilities, and as we successively drew white chips as opposed to red chips, the values came down dramatically, but it took about twenty rounds before the bids flattened out.

So, here is a place where people have really large cognitive problems here. Very afraid of losing forty dollars, and it seems as though in early rounds they over-estimate the risk to themselves and overbid for insurance.

So, that is what happens in the real auction.

We also ran the auction, again it is a victory auction, but we did it totally hypothetically. We drew chips out of a bag. We sat them there for the same length of time, and the bid distributions were significantly different, and they did not improve with experience.

In other words, hypothetical experience in this case did not make the hypothetical bids converge to the actual bids over trials.

Now, does that answer part of your question?

ATTENDEE: What I was trying to say is taking this to the real auction, and then use those distributions as the distributions that cover fat tails.

DR. SCHULTZE: The useful thing, however, is that if I have a real auction, and I look at my distribution of bids, even in a real auction, especially if it is for something like a low probability insurance policy, I get a fat tail.

If I do it hypothetically, I get a much fatter tail. If I take the raw mean of the bids, I get about twice the value, or two and a half times the value in the hypothetical auction as I get in the real auction.

But if I take the geometric mean, or the logarithmic mean, I get the same predicted value.

So, I can eliminate the hypothetical error by taking a logarithmic mean. This suggests that hypothetical

measurement error may be logged normally distributed.

DR. SCHUMAN: Can I just add one more thing on the random response technique, because I think it brings out the objective verses subjective questions. That has been used mainly -- have I used heroin. Have I ever stolen a car? Those are things I know, and I may not want to admit in an interview, so using the random response technique which completely eliminates the interviewer, knowing what I said in a sense, if very useful.

In the case of something like the Alaska study, or the others that we are talking about, these are subjective responses to something that the person probably hasn't thought about a lot before, isn't clear about, and it is more does the person shape their own thinking, not just their answer, in the direction of what might please the interviewer.

DR. DESVOSGES: Paul can I add just one minor thing? I guess Jake, the question I would have would be is the ability to be able to construct a situation in a laboratory

where nonuse values were really going on.

I think the more concrete it is, the more use-value oriented it is, and I think that suggestion works very well. To the extent that you move to the other end of the spectrum, I don't know, but that would be my concern.

DR. SCHULTZE: I haven't quantified this. The geometric mean for a CV study is about half of the raw mean, and if you look at my listing of lab experiments that have compared hypothetical values to final auction bids, we find they are about twice as high.

So, at least for use commodities, with perhaps the exception of the tree murder experiment as it has come to be called, we are dealing with use studies.

It appears that this is a correction that researchers should consider and suggest that raw means may be about twice as high as actual values they would obtain in a real market setting.

DR. PORTNEY: One last, quick question.

ATTENDEE: Schwartz, University of Michigan. I was

advised by the discussion about the arrangement between the open method and closed method, and all we know about the impact, the behavioral. I was very surprised by the discussion in the Schultze paper which focused on the open ended question, over-estimates or under-estimates responses.

I would like to learn more about the relationship. Basically, how did you pick those to begin with.

My second question goes to --

DR. PORTNEY: Second and even more brief question. Go ahead.

ATTENDEE: Pertains to the issue of mail surveys, and I just want to emphasize that from a logical point of view there is one aspect of mail surveys that is very difficult to deal with, and that aspect is surprisingly mail surveys show much less pronounced --

In the CV that is not at all true.

DR. PORTNEY: Bill or Howard?

DR. SCHULTZE: I was reporting studies in the literature, and I can say that the studies in the literature

generally take a range of values. There is not simply one value to which people respond yes or no, but rather they take a plausible range.

For example, five dollars, twenty dollars, fifty dollars, a hundred dollars, and usually in pretesting then you want to get values high enough that you choke demand down to something like ten percent or less.

So, usually in pretesting you look at an array of numbers, and you want the highest number to be such that it has choked off most of the demand.

So, is that responsive to your question?

ATTENDEE: It is not in the sense that -- as far as I can see, given that your responses are close-ended, -- what I am asking is how close -- when you calculate over-estimations and under-estimations -- how close --

DR. SCHULTZE: I am sorry. I can't answer that. I have not done these studies. I was citing studies in the literature. Maybe Mary Jo Keeley could answer that.

DR. PORTNEY: I guess what I would like to do in the

interest of providing the same amount of time for the final panel that we have for the others, is ask if we could change on the fly, so let me ask Jordon Louviere, Alan Randall and Rod Cummings to come up here, and at the same time please join me in thanking these gentlemen.

(Applause.)

DR. PORTNEY: By way of introduction for this session, let me say that going back to the time of the NOAA panel, and certainly continuing on since that time, one of the interesting questions in the CV debate has been the following:

Some people would allege that willingness -- answers to willingness to pay questions will always over-estimate true willingness to pay.

Obviously, that is an issue that is hotly debated, but people who believe you will always have some amount of over-estimate then ask: Well, is there a way to consistently determine how much of an over-estimate we have, so that we can then scale down the willingness to pay estimates from CV

responses and come up with a true estimate based on this calibration factor.

That is what this panel is designed to talk about. Our first paper will be presented by Jordan Louviere. He is currently Professor of Marketing and also part fellow in the Department of Marketing at the University of Utah in Salt Lake City.

His research has involved marketing, but of course a lot of research has gone into looking at surveys designed to elicit peoples potential willingness to buy new products that would be introduced.

Clearly what we are interested in is the percentage of people who indicated they would buy this, how many actually did. So, it is a calibration question in marketing. Hence, I think he is uniquely qualified to bring a perspective outside of contingent valuation to this question of calibration.

Jordan, the floor is yours.

DR. LOUVIERE: Can you hear me in the back? I guess we

typically use overheads, so I will probably be the only one here to do that, so I beg your indulgence in that respect.

My background is in transportation research and marketing research. I know very little about economics, even though I work with a lot of economists. I find I learned something new. I didn't know every day that I do work with economists.

My research interest is primarily in the modeling of drug behavior. I have been at that for about twenty five years. My Ph.D at the University of Iowa focused on conjoint analysis, which was an emerging technique in mathematical psychology, marketing and transportation in the early '70s. That is the perspective I want to bring today, is to talk about particularly with reference to techniques in the general family, conjoint analysis, particularly those based on random utility theory, which are consistent with most of the kinds of utility functions that many of you want to estimate.

Let me begin by telling you a little story about how I

got interested and why. I spent a lot of time on external validity work. In the early '70s I got invited a lot to transportation conferences, and I was typically the only person getting up during decision-making studies, in which I developed utility functions for hypothetical stated preference studies.

After getting clobbered is probably the most polite way to put it by economists, basically dismissing everything I was doing under the heading that this had absolutely nothing to do with what anybody in the real market does, please go away.

(Laughter.)

Most people I think would have done a mid-career correction, but it only made me mad. So, I spent the last twenty years working primarily in external validity work. Not only because it made me mad, but because one has to in a certain scientific sense, one has to feel comfortable with what one is doing.

And in the end, after I had enough time to think about

it, a very relevant question. Is it in fact a case that anything we ask people in hypothetical situations has anything to do with what they do in the real markets?

So, that is what I am going to talk about today, particularly with reference to some very exciting work that has been developed over the last three or four years, particularly Ben-Akiva and Morikawa's group at MIT, and a number of my own colleagues.

So, to change my affiliation slightly for you, in four weeks I will be at the University of Sydney starting a new department. If you are in Australia in the next several years, feel free to drop by.

(Laughter.)

So, my purpose today is basically to give you a brief overview, and I mean extremely brief overview of the theory and methods that have been developed primarily in marketing and transportation, measuring all stated preferences.

Let's briefly review why we are about stated preferences at all, when we have very strong econometrics

with dealing with preference data. The primary reasons are that there are considerable limitations in many contexts, particularly forecast and demand for new products, or changes in existing products that don't incorporate what we now have.

So, for example it is frequently the case that I get called upon to estimate the demand of at least the market potential for new products and services.

Often, there are literally millions of dollars riding on these demand estimates. These are not things we take lightly.

At the present point in time, I have projects in six countries for Fortune 500 companies, Federal agencies, et cetera, all of which involve enormous sums of money that are potentially on the table if we don't get these numbers right.

So, we are highly motivated to get these numbers right.

So, here are some of the reasons why RP data did not

serve us in good stead often. At least one characteristic of a new product may be different when it is introduced.

For example, the first toothpaste that comes into the market with gel foam for tartar control, that attribute was not observed prior to that in the market, so how are we going to estimate what the effect of adding that feature to a product is going to be by total demand or switching patterns in the market with RP data. We basically made a lot of assumptions.

More commonly, products are introduced, characteristics or attributes have a range well beyond current market values.

For example, when Compaq introduced their highly functional lines of PCs at much, much lower prices than had previously been seen in the market, when you get the first 500 m.h. check. Likewise, we often are called upon to estimate the demand for totally new category of goods. A current good example of personal communication devices, and these are the ones that will follow you anywhere, we don't

really have any market data on those. We have to do something else.

But in addition, even when we have revealed preference data, there are limits with what we can do with the revealed preference data in any real markets.

For example, it is often the case in real markets that the variation in explanatory variables is very limited, and marketness is quite common because products and services tend to copy each other, so we often find real markets and get a very, very limited range of variation.

For example, prices between products that compete in the market, making it very difficult to get a good estimate.

Likewise, in the product categories, that approximate efficient frontiers, there are lots of those. I hope everybody here likes that, by the way. If they didn't approximate efficient frontiers, at least momentarily, I think there would be some reason to be concerned.

In those kind of markets, product attributes or characteristics tend to be highly correlated, often they are

highly negative.

This makes it difficult from the average standpoint to get reliable estimates. Often there is no obvious objective measures available. So, we need to estimate models involving latent variables that pick up measures for things like product quality that we don't really have, necessarily, a hard measure for.

And often in many markets the number of auctions that are available are fairly limited, making it hard to actually separate out any of the product characteristic effects on choice.

So, obviously, the particular products that people are discussing today, there is no RP data available, we have to turn to something like the constructive market.

So, it strikes me that if you buy those arguments that in fact academic and applied research, and not market valuation, might benefit from something we learned in marketing transportation about how to use SP methods, but presumably you are only interested in what I have to say if

these methods are, in fact, reliable and valid.

Now, those two definitions have been already over-defined today, so I won't bother to rehash that.

So, what I am going to do today is I am going to review two general ways that marketers and transportation researchers often forecast demand for new or emerging products.

I am going to discuss how such measures have been used and are used in fact to make forecasts of real markets, and then I am going to review some case studies.

There are three in my paper, but I brought a few others along because I wanted you to see those three.

If you want to discuss these areas in general with me tonight or tomorrow, I can discuss many, many more cases beyond the ones that are in the paper.

So, the point I am going to make today, even though I know a lot of people aren't going to want to hear this message, is I think it is fairly unlikely that anytime soon in any case we are going to find a single number that we can

use to scale down, scale up, or otherwise scale CV or related kinds of willingness to pay or willingness to accept.

I think you will see why I believe that as I go through the talk today, so I will leave that as a teaser for you.

But first, we have lots, and lots and lots of evidence in the marketing literature, where this question is extremely important, and has been looked at by many people.

We have much evidence that suggests strong monotonic links between stated preference measures. These are things like asking people their intention to purchase a particular kind of product or category of products, or even a particular product, or take a particular vacation, or do a particular thing, on rating scales.

Notice, I said these relationships are monotonic. My colleagues in transportation dismissed most of the work in this field in the '70s, because they believed that they could actually go out and ask people whether or not they would use a new transit system if it was introduced, and lo

and behold, guess what?

Everyone here knows the answer, right? When seventy percent of the people said they would use the new bus system, transportation planners build a new bus system. How many people do you think actually used it? Ten or fifteen percent, and after enough of these, with a lot of money spent, and assuming they believed these numbers, the transportation planner dismissed this entire area.

Perhaps for good reason in the '70s, but as I go through the talk, I think what you will see is that we consider this extremely premature, and the question is not whether in fact seventy percent of the people who say they are going to do something will do it, but how can you transform the number seventy percent into an approximate percent that, in fact, will do it.

So, the two general areas of marketing -- I should say the two general approaches for measuring model preference for new products, the first is -- I am sorry to use marketing jargon, but if you are going to ever read our

literature, it is good to at least know something about what we have to say.

The first what we call prior concept test, this is almost exactly analogous to what people have been talking about today, whether it is referenda-type CVM questions, or open-ended willingness to pay questions.

What one has to do is to develop a concept statement; that is, to describe in great, glorified, boring detail to people what this actual new product or service looks like. And, the accuracy of the forecast that you are going to make depends absolutely critically on the comprehensiveness and accuracy of this particular description.

If you would like to see some of the ways in which you do this, let me recommend you to an excellent source. Glenn Urban and John Hauser's 1993, just published, book on Design and Marketing of New Products. I think you will be struck, particularly as you read through the chapters on concept, how similar this is to the problems you are interested in in CVM.

So, once we have a concept and we provide an elaboration of that concept, there are two ways to forecast based on concepts. The first is what we call norms-based forecasting.

Norms-based forecasting is normally confined to very large, well entrenched market research firms who do this a great deal. That is, over time, for many, many types of products and services, they develop empirical relationships between how people rate intentions to buy or purchase or use, or whatever, on rating scales and follow-ups in the market at various stages in the product life cycle as to how many people or what proportion of the market or what's the total number of units in the market related to those particular rating scales.

Most of these functions, by the way, for those of you who like discreet choice models, almost all of these functions look extremely logistic. So, it would be quite easy, if one had access to this data, to develop forecasting models based on typical logistical regression procedures

that should forecast the likely response of any particular population you are interested in if you had enough data or enough different broad categories over time.

The second way to do this was actually developed by Glenn Urban, one of the co-authors of the book I just mentioned. And, it is what we call "simulated laboratory test markets."

I will make it simple for you. We corner some of you guys in the mall. After we corner you in the mall, we invite you to an entertainment viewing experience.

We give you some short entertainment clips. We also embed a whole bunch of ads in these clips, just as you would find if you watched a normal t.v. program, some of these ads randomly placed depending on who comes in where and what malls are for the new product.

After you go through this, we typically run you into a simulated store which looks like a store with shelves, the whole nine yards. And, here are the products, including a new product, all listed at what we expect their retail

prices to be or the retail prices that they are offered at now.

We give you a set amount of money that should allow you to buy any one of these products. And, you have the choice to buy a product or keep the money.

If you buy a product and it's less than the total, we give you the money back. We use the estimate of how many people buy the product as an estimate of trial.

There are two components that are very important in marketing. We need to estimate trial and repeat, because we need to know where you are going to be on the adoption curve which looks fairly logistic until part of the categories get mature at which point the curves go down.

Now, to get the repeat rate, we follow this up at some period of time which approximates the end of purchase time, is what we call it in marketing. That is, we know about how frequently people use and buy products.

So, we redo the whole thing essentially by re-contacting people, offering the products again. And, we

find out from how many people we buy, we get an estimate of the repurchase rate.

I can tell you, from the published materials that are available on this, the models that employ this procedure are ludicrously accurate at forecasting introductory trial and repeat. I'm talking error rates on introduction of around plus or minus five to 10 percent.

So, it's good enough for companies to literally invest 60 to 100 million dollars at these kinds of estimates.

Now, there are a number of limitations in both of these procedures. And, these limitations also apply to all of the CVM studies that I've heard discussed today and that I've read about since my education in CVM began, thanks to Richard Carson, a number of months ago.

I'm not sure I want to thank Richard for getting me into this --

(Laughter.)

DR. LOUVIERE: -- but, anyway, any inaccuracy in the description that you generate for people cannot be accounted

for in any model after the fact. So, if, in fact, 25 thousand sea otters got killed in the Exxon Valdez spill, when you tell people only 2,500 got killed there is literally no way to readjust the forecast afterwards to correct for the fact that you were an idiot.

(Laughter.)

DR. LOUVIERE: So, since everybody likes to coin phrases today about phenomena, I would call this researcher's stupidity.

(Laughter.)

DR. LOUVIERE: So, all of us have attacks of stupidity to various degrees or other, but in order to control the error that can result from researcher's stupidity, obviously the more time and effort which is involved up front in doing this right, the pre-testing, the final testing, et cetera, the more you will reduce the area in the forecast due to this kind of inaccuracy.

The second limitation, which I consider fundamental to everything I've heard today, is that you can't identify the

effects on any of these measures due to particular characteristics of the products. I find it difficult to believe that the people in this room don't actually want to know the value of a duck or the value of an otter or the value of whatever.

For, when you give the value of one big scenario, perhaps two, it is impossible to separately identify each of the components of that description and place a dollar value on that. Surely, you want to do that.

I mean, right? I certainly want to do that. So, that brings me to what I am going to talk about for the rest of my time, which is basically what marketers and psychologists call conjoint analysis.

Many of the people in this room have probably been exposed to conjoint analysis in the past. I will refer to that as traditional conjoint analysis.

That's a special case of what I'm going to talk about, which is the much more general case of design analysis of discreet choices which are consistent with random utility

theory. For those of you that were exposed, at some point, to traditional conjoint analysis in the past, found that you didn't like it for whatever reason or thought that the models that you could get would not necessarily be consistent with economic theory, you were right.

The models that we are going to talk about today, in fact, are consistent with random utility theory, are consistent with the normal kinds of expectations you have for economic outcomes.

So, what conjoint analysis does, in general, and experimental choice analysis, in particular, is it recognizes that each product or, in this case, each particular environmental resource or event that you are interested in is really just one of a continuum of many, many, many possible environmental or resource events that share in common the characteristics that would be held by all such events. So, we marketers refer to these as positions.

If you think of this in a Lancastrian sense as

attributes, what we are really saying is that all of these phenomena are multi-attribute phenomena. If we know what the levels or values of the attributes are, we can completely characterize any particular event.

So, the Exxon Valdez is then not a unique event. It's just one of many, many, many possible oil spills that could have occurred; the Puerto Rico, San Juan spill being another example.

So, your objective in conjoint analysis from a marketing perspective -- this event will not necessarily be your objective, of course -- is to identify promising positions, that is promising bundles of attributes, promising combinations of the attributes, from a marketers standpoint to measure preferences for these bundles of characteristics, as well as the characteristics themselves, and then forecast a likely market performance. And, the newer approaches, most of which have been developed by many of my research colleagues, directly model the effect -- measure and model the effects of competitive activity.

Traditional conjoint analysis effect does not take competition into account except with some very strong assumptions, most of which are unlikely to be true in real markets.

So, in general, in experimental choice analysis, we view parts as bundles of characteristics that are valued by consumers for the benefits they offer, the problems they solve, et cetera. So, as I've said, any oil spill can be one of many, many possible spills.

And, in principle, the characteristics that describe any spill can be inventoried and a mutually exclusive and exhaustive list established, at least of the ones that matter. I mean, there's an infinite number of possible attributes of oil spills.

I don't care about all of them. I only care about identifying as many possible ones that actually matter to real people who are going to make these kinds of judgments and evaluations. And, I guess I can, which is where up-front homework comes into play.

Let me quickly describe two projects I am involved in right now to give you some flavor for this and for why all of you involved in these kinds of resource evaluations should care about these kinds of approaches. Right now, in northern Ontario, we have two ministries. One is responsible for timbering and logging and other things; and, one is responsible for tourism and recreation.

The loggers are busy logging. They don't talk to the recreation people.

We have a multi-million dollar remote tourism industry in northern Ontario where people fly into lakes to be left alone at an outpost or a camp or a lodge for a week or two and to enjoy the "wilderness experience." Now, I want you to imagine you are flying in on a float plane to this remote lake that you've just paid three thousand dollars to visit for a week.

And, suddenly you realize, as you are getting closer, that this lake is basically in the middle of a forest donut that might be anywhere from a few yards to a few hundred

yards. And, everywhere around it, as far as you can see, the area is clearcut.

(Laughter.)

DR. LOUVIERE: Once you get out to the lake, you can hear the logging trucks in the background. Is this still a remote fishing experience? Probably not.

(Laughter.)

DR. LOUVIERE: So, the Ministry of Natural Resources in Ontario has asked us to try to estimate the effects of these kinds of logging activities on the demand for remote fishing in northern Ontario. There is not one logging activity. There is a continuum of possible logging activities.

And, we've got to be able to take into account as many of them as possible to get a decent estimate of the likely effect of these things on demand. These things are also visual.

We have to present these things in the form of highly accurate visual images. Just the imagery required to do this for this study, which is all done using extremely high

tech multi-media graphics, costs 55 thousand U.S. dollars just to produce the booklets.

They are all computer-generated. And, I can guarantee you that none of you, if you saw them, would have any idea that they were not real lakes.

The technology to do this is here. It has been here. There is no reason for anybody in this room doing studies in the natural environment to not be making use of this extremely high speed, accurate way to depict changes in natural environments.

I know some of you are. But, I also know some of you are not.

A second, and very critical, question, not only here but in many other countries -- and I've actually worked on one project in the entire State of Oregon and another one for the country of Australia -- is what are the likely effects of congestion, pricing or other congestion relief policies going to be on the demand for future travel. Transport planners in the past have viewed this as a

relatively simple exercise. That is, you put a congestion price on a road and you find out what routes people choose or what modes they choose.

But, let me call to your attention that this is just a first order effect. If we put a congestion price on the artery out here around Dulles, not only are people possibly going to change routes, possibly going to change modes, but what else could they do?

I mean, there are a plethora of coping strategies people could enter into to avoid or mitigate the effects of this kind of policy. They could change their residence.

They could ask for a telecommute option. They could trade in their vehicle.

And, I submit to you, any model that doesn't try to take into account these second, third, fourth and fifth order effects will not forecast them. These are not simple projects or models.

I was going to put this up, but it's going to take me too long to go through it. And, I do discuss it in the

paper.

This is the general conceptual background that we have relied on for years in the particular paradigm I work in. It's not unique to me.

It goes back, I'm sure, earlier than this but as early as the 1950s in psychology. It's nothing unique here.

Now, on to random utility theory. Random utility theory isn't new either.

L.L. Thurstone, in psychology, first proposed random utility theory as a model for paired comparisons. That is, a model for the process that humans would use to form dominance judgments when I give you two things.

So, I give you Product A and I give you Product B. And, Thurstone proposed a model about how, in fact, you would decide to choose in repeated trials A over B.

This was extended by Dan McFadden in 1974; actually earlier than that, but the first usual citation is in 1974, the multiple choice case. And, since that time, it has been extended by many other people to allow us to incorporate all

sorts of interesting variations in choice behavior.

So, random utility view is, from many of our perspectives, that is as an outsider looking in and trying to figure out what real humans are doing, choices varied over occasions. And, we found out why. It's our job to explain that.

And, this variation can be explained basically by a systematic component. This is the component where we put into the model everything that we can think of, either that we designed to vary or that we think is a co-vary that explains differences in individuals, that I will get to in a minute, and then we add a random component that admits that we don't know what is going on or, at least, not as much as we would like to know.

So, as soon as we admit that there is a random component, then we are in a stochastic world. And, what we are interested in doing, at least from the marketing standpoint, is predicting probability that some alternative, I, in some set of competing alternatives, C, will get

chosen. And, that probability can be expressed as the probability of the systematic and random component of utility for Product I is larger than all of the systematic and random components for all the competing products.

So, the trick, you see, is to get as much into this systematic or explainable component as possible to reduce the size of the error component. And, that's what I do. I mean, my whole life is designing experiments that are consistent with this particular approach.

So, once you've got the variables identified, then we use theory and, in all honesty, usually lots of educated guesses or empirical evidence when that's available to specify decision rules -- what all of us, I think, will happily call utility functions, what our friends in psychology would call rules that people use to make the choices that they make.

Specifically, we use linear in the parameters and variables specification rules because they are great, great flexibility and generality. But, they also have

restrictions. They also have limitations.

They are not particularly good unless you have a lot of time and money and resources that fully align you to explore the underlying types of processes that my friends in psychology would be interested in. So, they are typically, at best, approximations.

In order to accommodate the differences in individuals that lots of you like to account for, particular income differences, et cetera, we introduce another component, which is just -- as an experimentalist or a statistician would call it, a co-variant term into the models to allow us to pick up differences in individuals. And, the consumers problem is to optimize that equation.

So, they are maximizing utility, which ought to make a lot of people in here happy and some people unhappy. And, different probabilistic choice models will allow us, depending on what assumptions we make about the random component -- for example, first don't assume the normal.

The distribution of McFadden wasn't even available to

Thurstone. And, it's interesting to speculate, as many people have done, that had Gumbel, the extreme value type one, if that distribution had been available to Thurstone he likely would have picked it which would have made psychometrics a very different field than it is today.

And, of course, this is consistent with Mike Hanemann's and other work in CVM that is based on a random utility theory.

So, the idea then is to design sets of profiles and to design sets of choice sets into which to place the profiles. Those of you that are familiar with conjoints stopped at Stage One. That is, you designed profiles.

If you want to study choice behavior, it's not enough to design profiles. You also have to design the sets in which the profiles will compete with one another.

And, you have to design those sets to be consistent with these stochastic choice models. That is a very hard problem and not a problem for which we have a general solution for, although we have some very, very good

solutions at the present point in time.

By playing certain tricks, one can use standard theory from the design of factorial and fractional factorial experiments to design these kinds of studies. If you are not up on experimental design, there is going to be a very, very large learning curve for any of you to try to arrive at this particular approach.

In any case, in 1983, George Woodworth, who is a mathematical statistician at the University of Iowa, and I introduced this particular approach where we proposed a whole set of particular ways to design choice experiments that were consistent with stochastic, multiple choice models. The problem was that up until we came up with these ways to design choice experiments that were consistent with these models and allowed us to test properties, there were cross experiments in psychology, statistics, economics, marketing and other fields. It was the paired comparison experiment.

The problem is, as we have shown in this paper and

we've shown in many numerical simulations since, if you want to study multiple choice behaviors, virtually the worst experiment you can pick to do that is the paired comparison experiments. It's extremely statistically inefficient if you want to study multiple outcomes.

The key for many people in here is to be able to design these experiments, such that you can test certain properties that violate the IID, that is independent and identical distributed, error assumptions of these models. Violations of those error assumptions lead to what are called violations of IIA, the independence of irrelevant alternatives, which is well known to many economists and psychologists.

Without being able to embed that in a model, I can assure you the forecast of models that do not take violations of IIA into account are terrible. So, this is a sine qua non if you are going to forecast well in the real markets.

In the model that we use, which is not particularly

theoretically indistinct but it has great application potential, is the Mother Logit or sometimes called the Universal Logit model as developed by Dan McFadden in 1975. All other logit models or members of the logistic regression family are nested under one of the logits. So, you can use Mother Logit as the general vehicle for testing any subset of functions.

Okay. So, what it comes down to is that you don't see this in any of the published papers, but embedded in all logistic models or members of the family of logistic regression functions is a scale of location parameters.

The error distribution is a three-parameter distribution. There is a mean, a variance and a scale of location parameter.

The scale of location parameter is what actually sets the scale for the utilities that all of you want to measure. So, that scale of location parameter is inversely proportional to the magnitude of the random component.

So, if you are not explaining very much about choice

behavior in any market, real or constructed, then the utilities are going to be very small. And, you are likely to conclude that many things are not significant.

If you are explaining lots about the behavior in any particular real or constructed market, then the utilities will be very large. And, you are likely to conclude that many things are significant.

What was proposed by Mosha Ben-Akiva at MIT, a Ph.D. student who worked with Dan McFadden four years ago, was that if we had real market data and we had a constructed market, then a very simple relationship exists between the scale parameters in the real market and the scale parameters in the stated preference market. And, that is, they should be in a simple ratio to one another.

And, if were able to estimate the same utilities from the two different sets of data, they should be proportional. Now, one can obviously test that as a restriction in a statistical test. And, one can estimate the value of the ratio of the real market to the stated market.

And, what showed at this conference that I put on four years ago was that surprisingly, in a set of transport data in the Netherlands, they were able to retain the hypothesis that the two sets of utilities were proportional up to rescaling by this positive constant. Well, I had so many sets of revealed preferences, stated preference, data available to me, I can't even tell you.

So, I immediately ran home and started seeing if this was true in all of my sets of data. And, I'm here to tell you that in virtually every single set of data that we have examined since 1991, this restriction has been shown to hold.

And, I'm here to show you two or three sets of data. How much time do I have?

DR. PORTNEY: Twelve minutes.

DR. LOUVIERE: I don't know how many of you have seen this paper, the Adamowicz, Louviere and Williams paper that just appeared -- at least, I think it just appeared. I've been told that it has appeared.

But, I would like to quickly go through this study for you, which was the first one that we are aware of in your area that actually tests whether or not the parameters of a revealed preference, a stated preference, model were proportional to one another. So, the context is a stream improvement on a major tributary to the Bow River in southern Alberta. Many of you may know that the Bow is one of the great blue ribbon trout waters in North America and the source of millions of dollars of tourist revenue to the Province of Alberta.

So, the question was, when we introduced this stream improvement, basically what was the value of doing that. So, we did this in two ways.

One was a traditional travel cost survey in which we asked people basically where they had been -- where they went last for forms of water-based recreation activity. And, after we asked them that, we recruited them to participate in a follow-up mail survey.

And, over my protest, a particular survey research firm

in Alberta was hired to do that survey. The result of that, for all of you who may have read the JEEM

ere were no individual

differences in our models, was that the survey research firm totally screwed up the data. And, we were unable to match any of the individual differences from the phone survey or the mail survey.

That's not in the paper. But, that's what happened.

(Laughter.)

DR. LOUVIERE: John Puswade is a student of mine who just joined the faculty at the University of Florida to get his Ph.D. under McFadden. We proposed a simple way to use existing multi logit software to actually estimate this scale ratio and test it.

And, the idea is as follows: You measure and code all the attributes that are common in both data sets if there are only two data sets. If there are effects that are unique to either data set -- for example, alternative specific interests will be unique to different data sets,

because you are going to use them to satisfy the site conditions in the logit model.

But, if you've got outcomes which are real streams in a travel cost model and in a stated preference model you will see people only choosing between pairs of things, when we've got a bunch of alternative specific constants for real places and you've got a bunch of alternative specific constants for real places throughout the cross model and you've only got potentially two alternative specific constants in a stated preference model, there is no reason to expect that those have any necessary relationship to one another. It's only the utilities that matter.

So, you can captivate both data sets into one large data set. You relate the data set that you want to test the hypothesis on by multiplying it.

And, essentially, you did a grid search by starting with what you think is a reasonable multiple. You multiply the design matrix, in this case the SP design matrix, by some number that you think will rescale it to the RP data.

You can show under general conditions that this procedure will converge very nicely to a unique maximum. And, if that unique maximum that the data set you are trying to rescale fits -- this is the first data set -- as well as it's going to fit, then you test whether or not that restriction, the difference in all likelihood between the two separate models and the joint models for a statistical likelihood.

So, that's the idea. For those of you that are good at programming, a thermal estimation will be much more efficient than the procedure that we propose.

Okay. Well, here is what this Bow River study looked like.

Here is a set of instructions. But, prior to receiving this, people were asked to evaluate actual real streams and lakes in southern Alberta. Are the attributes varied in this next task?

So, people already know what the levels of the attributes are. And, they know what the numerical or

qualitative values of them are. And, they know how they apply to real streams.

So, that's the question. Here is what we gave them.

So, each person got, from memory, 16 of these. So, the person's task was -- and I suppose this addresses the question that Professor Schumann raised about "don't knows." We don't use "don't know," but we sure allow people to do what they do in real markets, namely choose not to choose.

We happen to think that's very important. And, it has very important strategic implications from our view.

So, a person basically can choose a standing water option, a running water option or any other non-water base recreation option or stay at home. And, we asked them a series of questions.

So, we are modeling basically. Each person got 16 of these. There were four sets, four versions of 16 things, for a total of 64 choice sets designed according to -- if you want to know more, read the JEEM paper.

But, the design was basically two to the fifth by four

to the fifth by two to the fourth by four to the fifth. The standing water options on the left were the two to the fifth by four to the fifth crossed with the running water options which were two to the fourth by four to the fifth.

As I've shown in a number of my papers, you merely have to select the smallest main effects plan from this particular pool factorial to actually develop the choice sets that you want to present to people. And, that design will have the property that you can test IIA as well as lots of other fun things.

Okay. So, there are actually two to the ninth by four to the tenth possible combinations of attributes that define pairs of these water options.

This allows us to estimate all generic parameter gains. But, the price is the same regardless of the option. Or, the price coefficient differs whether it's standing water or running water.

We can also estimate and test that restriction against an alternative more general restriction that the parameters

are alternative specific. We can also get a complete test of IIA, although that's not presented in the JEEM study.

And, respondents were randomly assigned to each of these versions. Okay.

Just to cut to the bottom line, the JEEM study will show you we were able to retain the hypothesis of parameter per proportionality in that study. And, as an interesting aside, in the travel cost component of this, the travel cost model, about three-fourths of the parameters were not identifiable because the characteristics of the alternatives were perfectly linearly related to the alternative specific constants of the destinations.

As soon as you combine the two data sets, however, you now reduce the collinearity and now we can estimate all of the parameters in both the RP and SP data, an advantage that you might want to consider.

Here is a second study. Now, this isn't really freight shipper choices. I can't tell you what it is, but it's close to shipping freight.

So, this study was done in three major North American cities. We had eight competitors who vied for people's business in this shipping business.

They fell actually into two groups by the form of the service that they provided. So, what we did was to design another one of these collective factorials, a four to the fourth by two to the twelfth.

And, we embedded in the sets that we showed people a third option, which is always available in every city, that people could use that represented another form of service. And, as in the past, in the previous study, there were 64 pairs. And, everybody who participated in the study answered 16 of these sets.

The people who participated were people who actually made the freight hauling decisions for their companies. So, here is what this looks like.

Once again, prior to this, is a fairly significant number of pages in which people evaluated real shippers on all of the attributes that were varied in the experiment.

So, they were quite familiar. It was also included a glossary of pictures and all kinds of other stuff explaining what all of the terms were.

Okay. So, here is what this looked like. You will note it's pretty similar to the previous one.

And, people had to decide whether or not they were going to seriously -- this was done to test something in marketing called consideration sets. Many of you who may have run utility models would recognize this as trying to pick up something in the order of choice sets.

That doesn't work, by the way. But, moving right along. So, what we are really modeling in this Swait, Louviere and Williams paper is the choice of which particular option.

We randomly sampled 200 people in each city. We made sure the choices. We asked them what they had done last and what they usually did, et cetera.

Our response rate was around 50 percent. And, when checked against the norm of statistics, census statistics,

for those cities, there was virtually no differences between our sample and the city sample even though we have the concern about non-response in this case.

So, we did -- this is a paper that is going to come out in a special issue on Stated Preference Models and Transportation, edited by David Hincher, later this year. We have developed a sequential approach to estimating this model.

We first -- well, we adopted the assumption that the utilities that were most reliable are not the RP utilities, they were the SP utilities. So, we first fixed the RP utilities at their SP values.

And, then we rescaled the attribute columns of the RP we fixed to conform to SP units. Then, we allowed these alternative specific constants to vary free in the two sets.

And, then we rescaled -- after we estimated that model, we then sequentially rescaled the RP design matrix to SP units so that we could take advantage of the superior statistical conditioning of the SP units.

We then tested parameter proportionality for these data. And, this was rejected big time.

So, we could not confirm the hypothesis of the two sets of utilities in the same way that Ben-Akiva and Morikawa had proposed were, in fact, proportional. But, what we noticed was there were only a couple that were off.

And, we also noticed that the rescaled model seem to predict extremely well. And, there were little differences in the log-likelihood.

So, sure enough, when we went snooping further, we found that the rescaled model, which was primarily based on the stated preference, predicted virtually identically to an RP model, estimated only from the RP data. And, this may surprise some of you, but this is not a unique finding.

We frequently find that models estimated from RP data outperform -- from SP data outperform the RP models that are actually estimated from the RP data.

And, something we have just completed, sponsored by the U.S. Forest Service. This is an explicit attempt to compare

external validity using totally independent samples of people in which we take different measurements.

So, here are the following conditions we looked at: One sample of people, we only asked them where they went skiing last. We asked them some stuff about their skiing habits. And, we collected demographic information consistent with the U.S. Census of population.

The second condition, we asked exactly the same things but we gave them a choice experiment that I will show you in a minute.

And, the third condition, we asked them the same things and we gave them a totally different stated preference test based on a new random utility model I developed a couple of years ago called the "Best/Worse" model. And, I will show you what that means.

Here are the response rates for those surveys. Here is what the ski area choice task looked like. Once again, this comes after people have made a whole bunch of judgments and evaluations about 10 actual ski areas on the Wasatch Front

in Utah.

So, this is going to look just like the other things you saw. So, I will move along in the interest of time.

And, where is what people basically were asked to do. They were given two ski areas.

And, they were asked which one they would choose or they could choose to do something else if not ski at all. That is, they had the choice not to choose.

Here is the second SP task. And, in this particular task, we are simply going to ask people to tell us which attributes are respectively the most attractive and the least attractive and then make a judgment, very much like a referendum, as to whether or not you would actually use the ski area if it was available.

So, here is what the task looks like. So, a person has to literally explicitly trade off the values of the attributes, one for the other, in this task.

They make the explicit pool judgment of which one is the most attractive, which one is the least attractive.

And, then we ask them whether or not they would use it.

This allows us to pick up the demand effect. And, we are going to jointly estimate the model together.

There is a very special model that underlies this. And, I'm not going to go into that.

This was done, by the way, in conjunction with the Survey Research Center at the University of Utah. We got virtually identical response rates in each of these.

We are not concerned in the least about nonresponse, because this is a test across these three conditions. So, what we are concerned about is whether or not there were any differences in these three major sub-samples.

And, several hundred comparisons across the variable have been measured in the survey. Only two of these comparisons are statistically significant. Therefore, there are virtually no differences between any of these sub-populations.

So, here is the primary result. How much time do I have?

DR. PORTNEY: Negative two minutes.

DR. LOUVIERE: The clock is still ticking. These are the two scale ratios.

Notice that the choice experiment produces a scale ratio of around point 4. The best-worse or maximum difference task produces a scale ratio of about 28.3.

One of the reasons I told you I doubt a single scale constant to rescale CV data is going to be found is because it's going to depend on the technique you use as well as the kind of product category. But, here's a picture of what the actual test looks like.

Now, in addition to those two measures, I'm going to graph on this one graph. We collected a ton of stated preference measures in this study.

And, since I wrote this paper, we've estimated all of the models on all of these stated preference measures. And, the only thing going on in all of these models is differences in variability.

That translates into differences in reliability. That

is, if every single one of these stated preference techniques measures the same fundamental underlying data generation process, all of them should apply to a series of positively sloped straight lines through the origin.

Hopefully, you will agree visually this comes really close. And, I didn't force the functions to do that.

Jarvis Swait and I have just completed the statistical test which requires a FIML estimation to jointly rescale all of these together and test the hypothesis as to whether or not, in fact, they all rescale. And, they do.

So, we cannot reject the hypothesis that every single one of these measures meets the same underlying utility measures. So, here's the conclusion.

These measures are equally valid but unequally reliable. Therefore, what you would want to do is now undertake research to try to figure out which of these methods is more reliable in which kinds of situations than others.

So that we cannot begin to understand whether we prefer

to use one as opposed to the other. And, those considerations bear lots of things.

Some are easier subjects to do. Some provide more interesting and reliable information.

And, I have nothing to say about that at the present time, because the field is simply too new. And, I think I will shut up.

DR. PORTNEY: George, thank you. We have two discussants, Alan Randall of Ohio State University, and Ron Cummings.

I will save Ron a little time, because he's the editor of JEEM. He's going to want you to subscribe.

But, we will hear first from Alan Randall, a Professor in the Department of Agricultural Economics at Ohio State. Alan has done a lot of the original work in bidding games, has done contingent valuation experiments on visibility improvements, among other things.

And, I am happy to turn the floor over to him for 15 succinct moments.

DR. RANDALL: Oh, so much to say and so little time. I had prepared 20 minutes of commentary and, of course, during the course of the day have prepared a couple of dozen things that seemed worthy of a sentence or two.

But, on the other hand, some of the things that I had planned to say have been given airing as well. So, certainly we can have some fun here and stay fairly close to the time limit.

Firstly, back in pre-history, roughly, say, about 1950, a couple of parallel suggestions arose with respect to evaluating non-market goods. That is, we could learn something by seeking self-reports. That is, by asking people what it's worth to them.

Or, we could learn something by examining and analyzing and manipulating market data in goods believed to be complements, substitutes or whatever. And, we eventually, I guess, settled on complementarity to get started.

And, as Mike Hanemann pointed out, already by 1965 we reached the point when the first comparison between

techniques of both kinds in the same valuation task was published. And, that seemed to sort of set the agenda for the next generation roughly.

That is that external validation was obviously required for both the contingent or stated preference, or whatever techniques and for the -- what I would, I think in all honesty, refer to as the conceal as opposed to reveal preference methods; that both required external validation. And, we could do this in various ways. But, they mostly involved comparing the results of one to the other.

Over that roughly one human generation, enough studies were done that Richard Carson and his research associates were able to find and tabulate about 80 published studies, making a total of more than 500 specific comparisons. And, so the agenda was external validation by comparison.

Then, there was an abrupt paradigm shift promoted by, I think frankly, a vocal minority that were, I think it's fair to say, encouraged to pay attention to this issue suddenly, precipitously and in the heat of actual potential

litigation. And, the attempted paradigm shift was from external validation to calibration, the notion being that of these various ways of knowing about values, some kinds more suspect than others, and that the more suspect kind needed to be adjusted.

And, the recommendations seemed to -- to the extent there were recommendations, they seemed to be aligning up that the adjustment needed to be downward. Now, that seemed to me to be an attempt at paradigm shift but one that I don't think is totally established.

Now, let's go back and take a look at the comparison studies. Michael, again, summarized those.

The basic result from the meta-analysis by Carson et al. of these 540-some comparisons and from studies is that the mean ratio of contingent values versus values from revealed preference techniques are some on the order of point 85, point nine. The confidence intervals are fairly narrow.

And, it seems like what we need to roughly do, if we

are going to do calibration, is to calibrate contingent values upward by about 10 or 15 percent. But, I don't put an awful lot of weight on the notion that we can learn an awful lot from specific comparisons.

I think we can learn more from this meta-analysis. And, there is a weight of evidence and a kind of gravity that is given to this by the fact that many studies are involved.

But, nevertheless, there are limits because the comparisons are necessarily of apples to oranges. And, there is some limit as to what we can now learn from that.

One of the comments that I must make is that many of the studies cited by partisans in the recent discussion as indicating one way or another the need for calibration are, in fact, studies which were included in the Carson meta-analysis. So, that in some ways that boils down to an argument of a sample with a good deal of selectivity in the choice of the samples.

I want to comment briefly on some problems with some

kinds of comparisons that are receiving some current discussion. I guess, to start out, it's not really a very interesting question to ask does everybody always tell the truth about everything. It doesn't seem to be too much of an interesting question.

We are fairly sure that -- well, let's put it this way: The decision structures we have developed give little weight to self-reports of innocence in criminal proceedings. The collection of income taxes is done by self-reports but subject to audit procedures that are believed to be relatively fearsome.

There is an example that I observed just drinking my morning coffee on one of the morning talk shows a couple of weeks ago that was kind of interesting. The researcher was in to report about the phenomenon of sexual relationships of the workplace, which he hypothesized was undergoing change due to the changing composition of the work force.

And, so he reported his first result. That is that one question, 85 percent of people reported having had at least

one significant sexual encounter with someone at the workplace.

This is pushing kind of the limit here, in that I'm not that convinced that 85 percent of the population have had significant sexual encounters of any kind. But, it's getting close --

(Laughter.)

DR. RANDALL: Okay. Then, the follow-up question was: How many of the whole population has consummated the sex act at the workplace during working hours?

Okay. And, the answer was 65 percent.

(Laughter.)

DR. RANDALL: And, the moderator shared my skepticism and said, "Come on, you know." And, the researcher reported, breathlessly, "They gave more detail than that." Forty-five percent of them, when asked where, reported that it was on the boss' desk.

Okay. There are some responses you can't believe.

(Laughter.)

DR. RANDALL: So, the interesting question is: Can we construct questions that might generate responses that can be believed?

There is no very interesting reason to take seriously differences -- the question of whether voluntary contributions hypothetically and in the real generate the same result. And, having said that relieves me of worrying about the particular deficiencies of the Seip and Strand study.

The only comment I would make on that is that I have talked to Strand about this. And, he admits to being kind of surprised and a tad embarrassed that the study has gotten so much press.

The next question concerns hypothetical purchases of ordinary private good objects versus real ones. And, there clearly is a problem.

And, those who are doing those studies have recognized that there is no real commitment when it comes to, for example, a hypothetical purchase offer price, but perhaps

the real commitment, in the case of a binding offer price. And, you get the result that pretty much everybody could expect. That is, the difference between a hypothetical and actual behavior.

Now, the matter of referenda seems dramatic here. And, there are some claims about that contingent policy referenda have better incentive properties.

But, it's important, I think, that we say exactly what we are claiming here. Consider a hypothetical referendum.

And, I think a hypothetical referendum would have to go something like this: Here is a proposition at a price. I want you to tell me how you would vote on that under the condition that I promise never to tell anybody else how the vote went.

Okay. That seems to me to be a hypothetical referenda.

The referenda that I've done ad nauseam on economic issues like the funding of schools in the State of Ohio have the property that is being discussed here, that they occur in a certain place at a certain time, in a certain fashion

and manner and after some campaigning. Okay.

The contingent policy referendum belongs somewhere in a different category again. What happens is that we kind of know that there are several million vigorous and well-meaning public servants in the country who would love to provide projects and policies to the American public and to tack a price on to their taxes or the public debt.

Okay. And, so we approach people and say, "Authorities are trying to decide such and such an issue at such and such a price. And, they've asked me to go out and ask a sample of you, how do you vote."

Okay. That, we would argue, your response to that represents a kind of irrevocable commitment, in that you are not going to get a chance to change your mind. And, the response, if it's positive, will give aid and comfort to this army of bureaucrats who are eager to do good things for all of us.

I guess this is merely elaboration of a point that Richard Carson made. There is no claim that the civil

servants will then go out and do exactly what the vote suggests.

But, there is a reasonable presumption that the likelihood of policy implementation is monotonically increasing with the proportion of yes votes or the amount of stated willingness to pay. And, I think that kind of arrangement has, in fact, better incentive properties than does, say, hypothetical purchases of private goods.

Now, that's enough on comparisons. Well, I guess one more comment on comparison studies is that a lot of the estimation and re-estimation comparison and recomparison that we are seeing in the recent literature and discussion, lots of things seem to depend, the important results seem to depend, on the treatment of few extreme observations and the choice of statistical tests, which you may or may not be sensitive to extreme observations.

You get different results, that's true. And, if that represents a problem, then it seems a simple matter to promulgate guidelines on the handling of that and choice of

tests as they affect the sensitivity of results to extreme observations.

Now, the research agenda that has emerged in recent years that I think is significant and represents a really meaningful paradigm shift is the one that Professor Louviere's paper falls into. That is that we can learn some things from self-reported preference in values and we can learn some things from market-simulated goods.

But, both of these windows on preference are kind of translucent, at best, and distort it. And, there is a real possibility that we may dramatically enrich our learning if we are able to combine both kinds of data.

This is an agenda that I support and have supported publicly. It's nice to see the work with conjoint techniques and in related fields that are not necessarily immediately in the non-market business from Professor Louviere.

But, I point out that some of our own -- Ted McConnell, Trudie Cameron, some of us have been doing studies in the

same kind of spirit. And, I certainly support this research agenda and believe, quite frankly, that it is the wave of the future.

There is an implication for policy analysis and damage assessment from this. And, that is that we all, everybody who cares about improving our ability to learn how things are valued, has a strong interest in avoiding a situation in which guidelines are promulgated that serve, among things, to stifle innovation.

A couple more comments on Professor Louviere's paper. The scale factor is miraculously opaque. It has things to do with the variability in the data.

We don't read the scale factor of point so and so and say that means that we simply raise the mean or the median by this amount. It's not about that. It's about something else.

And, the message that I get from Professor Louviere's paper is the very strong message of, firstly, consistency between stated preference -- between the utilities implicit

in stated preference observations and the utilities implicit in actual behavioral observations, that these are consistent, that what we can learn from either one can be enriched by combining them, and that this business of simply putting a thumb on the scales to move mean or median valuations up or down receives no particular endorsement from the Professor's paper.

I want to finish up with a comment that has to do with rule-making. It seems to me that there is not a coincidence.

It's not just a coincidence that late in the NOAA rule-making process the business of calibrating, dividing contingent valuations by two in the absence of a better number, has -- that's kind of come out and been released to us. And, so I think a section on calibration at a conference like this, it's kind of implicitly asking for some comment on that kind of procedure.

And, the way I would like to look at it is in terms of what incentives does it give to those citizens, the

respondents, who are reporting the values to us. The point is that, while we have discussed calibration as though if it's done at all it will be done secretly back at the office after all the numbers are in, the respondents will never know and never understand what is going on.

In fact, if calibration is to be done, that will be announced in the Federal Register.

(Laughter.)

DR. RANDALL: Now, if we view contingent valuations as a kind of a principal agent gain -- the respondents are the agents -- they are providing value statements under implicit conditions, that the principal will listen to those and make project decisions. We will add up votes and decide. Or, we will add up willingness to pay and compare it with project costs or whatever.

We are going to announce in the Federal Register that in the future the principal is asking for willingness to pay statements or contingent vote statements subject to the rule that in the case of willingness to pay, the principal will

divide that number by two; or, in the case of a referendum vote, that the principal will apply that number to a tax price of one-half as high as the one stated in the questionnaire. Now, I would argue that the best of our respondents could solve that problem and if they can provide an individually optimal answer to the standard CV question, they can provide an answer to the new question. That is the one in which the calibration rule is announced.

But, I think not everybody will get it right. And, the data is likely to become how to interpret noisier and nastier.

But, what if some of the respondents say this might be a repeated gain? We seem to have bumped off the principal with our previous reports and led him to go on dividing things by two.

And, so once we have adjusted to that, he may still be unsatisfied. He might want another round of calibration. Okay.

And, so I can imagine this whole thing degenerating.

And, in the extreme, agent reports to to infinity and the principal eventually divides the sum of those by infinity and nothing is learned at all by the exercise.

Now, in the most recent thing that has been circulated, the Department of Interior has offered an even more terrifying suggestion. That is that after we have looked at what the agents report, the principal will try to figure out how to calibrate it. Okay.

So, the agent's task is to report to me your value under the condition that I will take a peek at it, move it about as I like and then decide what to do about this project. I submit to you that rules along that line could only have the effect of making contingent value statement or contingent referendum votes even less reliable than they currently are.

DR. PORTNEY: Alan, thank you very much. We will hear next from Ron Cummings.

Ron is a Professor of Economics at Georgia State University, editor of the Journal of Environmental Economics

and Management, and co-author of a study that has received a lot of attention so far today.

Ron.

DR. CUMMINGS: The last one. And, everybody says, "Please, be on time. Let me get out of here."

(Laughter.)

DR. CUMMINGS: I will try to keep this relatively short. I do want you to know that this is kind of loaded against you, that they purposefully put me on last so that everyone that attended the RFF lecture last week doesn't have to stay. So, if you see someone with their head on the table sleeping, they've already heard this.

I'm just going to limit my comments really to two sorts of things. I want to talk to you about the recent research that deals with the question of the accuracy or the validity of CV value under different institutions.

I will comment very briefly on what I think the implications of those are for further research. And, then if I haven't gone over my time, there is one thing that I

would like to talk about that deals with publication strategy.

(Laughter.)

DR. CUMMINGS: Now, I am going to be doing the kind of thing that my colleague, Alan, doesn't like. I'm sorry.

I am going to use experimental techniques to try to get at the question, really a very simple question, what is the relationship between amounts that the people say that they will pay and amounts that people actually dig into their pocket, put it on the table, how much they will actually pay. Okay. That's all I want to know.

Now, my inquiry into this question really adopts what is sort of the basic stance that underlies experimental economics. And, it's very simple.

And, it goes something like this: Take any general theory, take any general proposition, that you say is going to work out there in that complex world and if it's going to work out there in that complex world then it should work in any little special, simple set of circumstances. Okay.

If it's going to work out in the complex world, then it surely ought to work in a very simple world. Okay.

So, we have heard a lot and we have heard for years and years that one of the biggest problems with contingent valuation is, "Oh, my God, you know, we've got to...," you know, our colleague, Professor Louviere, took a great deal of time telling us, you know, the importance of describing the commodity, making sure they know what they are valuing, all these kinds of things. We ain't going to have that problem.

We are going to take very simple little goods. You can hold it in your hand. You can rub it against your tummy.

Whatever turns you on, you are going to understand what this thing is that you are going to buy. Okay.

Now, before open-ended experiments became, you know, the dregs -- for reasons that I'm still not sure that I understand. The first set of experiments that involved an open-ended experiment, it has already been discussed today in the recent issue of Land Economics. I won't spend much

time on it.

What we do is very simple. We take a simple painting by an unknown, take one group of subjects, ask them their maximum -- "Take this picture. Take it right home with you today. What's the maximum that you will pay for it?"

Take another group of subjects and we run a second price auction. All right.

And, that has already been discussed today. We know that, assuming the subjects understand their dominant strategy, that their dominant strategy is to tell the truth, and we run that experiment and got these results.

The average willingness to pay under the CVM was 37 dollars. The average of the second price auction was $9.50.

But, then we got a lot of flack. And, it started, I think, with Mary Jo Kiley. You know, called me up and said, "Ron, you are a dummy, because you can attribute the difference in those bids possibly to two things. One is the difference between real and hypothetical payment. Secondly, they are two different institutions."

And, if you believe that institutions count, then you don't have a compelling case but the difference is attributable solely to differences between real and hypothetical payment. So, we redo it.

Now, what we are going to do is we are going to -- if we going to redo it, we are going to have a CVM. We are going to take about 200 subjects, train them all in a second price auction, try to get them to understand the dominant strategy is to tell the truth, then take half of those subjects trained in victory auction and will not take a hypothetical regression. The others take part in the re-auction.

Now, what's the difference between the CVM and the hypothetical victory auction? They are both hypothetical. The difference is institutions. Okay.

The difference between those two and the real -- the difference between a hypothetical victory auction and the real victory auction is simply real or hypothetical payment.

So, if you find, as we do, that there is no significant

difference between the CVM and the hypothetical victory auction that is consistent with evidence that would allow you to reject the hypothesis that the difference in bids is attributable to the institution, okay, you can then focus on what is a significant difference in bids between the hypothetical victory auction or the CVM and the real victory auction.

Now, under normal conditions, you would say, "Well, now that is just one experiment. Gee, the results are interesting, but you ought to replicate and see do you continually get this result."

Well, about this time, we learned that really open-ended stuff isn't what ought to be done. If you want to get the truth, God love you, you've got to use dichotomous choice. Okay.

And, because dichotomous choice now is incentive compatible. Are you ready?

(Laughter.)

DR. CUMMINGS: Well, why is that? Well, you are going

to either say yes or no. You are going to tell the truth. Okay.

Well, now that is certainly a testable hypothesis. But, again, you use the principle, if the general principle is dichotomous choice is incentive compatible out there in the real world, in that real complex world, then, damn it, it ought to be compatible in very simple special cases.

We can test that hypothesis. We are going to take three goods.

We are going to take a very little simple solar-powered pocket calculator. Okay.

You are going to say to your group of subjects, "Here, if you could pay three dollars to put this to put this hummer in your pocket and off you go home, would you pay the three dollars?"

You take another group of subjects: "Will you pay three dollars to put that hummer in your pocket and get along home?" It's very simple. Okay.

Let's take another good. We will take a box of

chocolates or a box of chocolate truffles, a 15 dollar box of truffles, take one group of subjects, "You can take this sucker home with you for three bucks. Will you do it?"

All right. Another group of subjects. "Here's a box of chocolates. Will you pay three dollars to take them home?"

Well, now, in both of those experiments, we used student subjects. And, a lot of people got all excited, "You used student subjects. Everybody knows you can't trust students." All right.

(Laughter.)

DR. CUMMINGS: So, we are going to do another one where we are going to use an electric juicer. And, we are going to get people at church.

They have just praised the Lord. God love them. If they are ever going to tell the truth, they are going to tell it now.

(Laughter.)

DR. CUMMINGS: Bring those suckers out of church and we

are going to ask them, "If you could take this juicer home with you, you know, just go right home after church and make a little orange juice, eight bucks, would you do it?"

Now, here and actually in the chocolates, these last two, we are going to go and sample. We are going to ask those church-goers, "If you could take this, buy this sucker for eight dollars and take it home with you, would you do it?"

Then, we are going to tell them, "You are not stuck with your old answer. But, I mean, guess what, here it is and you can take it home with you. Will you pay your eight dollars?"

Now, you are going to have to do something else here because of a possible ordering effect. Since I've asked them the hypothetical question first, it could effect, you know, the real response that you get.

So, we will do another experiment and ask them only the real question, compare the distributions of the two real questions. If there is no difference, if that is at least

consistent with evidence, that would support a rejection of the hypothesis that you get an ordering effect.

Here are the results of that three set of experiments. At the top is the electric juicer. Those were the religious folks.

You know, actually Glenn -- this is the first time he has ever been in a U.S. church.

(Laughter.)

DR. CUMMINGS: He glowed. He had a halo for a while.

(Laughter.)

DR. CUMMINGS: You will see in the hypothetical you get a 41 percent response rate, 16 percent when it's real. Here is your out of sample test for an ordering effect. There is no significance difference between the two that is consistent with a rejection of the ordering effect hypothesis.

The chocolate truffles, the same thing. This is an out of sample test.

Again, under normal circumstances, you would say

-- we could probably sit around and come up with a lot of ways of doing this study. So, you know, we ought to replicate these experiments, get people to throw stones at you and do them again.

But, then about this time, there is a new truth. If you want to do this right, you don't replicate a market.

We are told that in the referendum, it is the voting format now that is incentive compatible. If respondents desire the program at the stated price, they must reveal their preferences and vote for the program.

Now, what we have is the notion if we appeal -- in the social choice literature, if you appeal to the simple democratic model, we know that the referendum format is incentive compatible if, one, there is only two items of choice and, two, if subjects feel that their utility is affected by the outcome of the referendum. All right.

Now, people in the social choice area are still battling about whether people in real elections feel that their utility is affected by the outcome of the referendum.

All right.

What about participation in a hypothetical referendum? People are going to feel that their utility is affected by the outcome of the hypothetical referendum.

Well, surely that is a testable hypothesis. We are in the process of testing just this hypothesis now.

And, let me just sort of very briefly tell you and I can give you some preliminary results. Write me in a couple of weeks and I will give you some more.

What we do is we take a set of subjects and we describe -- we describe this is really a nonuse good. But, this is a group of subjects and we describe a group of people that live in Albuquerque, New Mexico, that are primarily low-income Hispanic families that live in an area that overlies a contaminated aquifer, very poor people.

A lot of them don't know if for sure their water is contaminated or not. If it is, they don't know what to do about it. All right.

And, there is an environmental organization in

Albuquerque that is working with us. And, they are willing -- what they want to do is they want to produce and distribute a guide that would tell these people exactly where the area of contamination is, if their well is contaminated to tell them what to do and if it is contaminated what recourse they have. All right.

But, they don't have the money to produce this and distribute it. All right.

So, we had our people vote on this proposition. Everybody in the room is going to contribute 10 dollars to the Southwest Research and Information Center. That's the organization that is helping us on this. And, the contribution will be used for the purpose of preparing and distributing the Citizen's Guide to (N) Households.

The rules of the referendum -- how am I doing, Paul? Very briefly --

(Laughter.)

DR. CUMMINGS: If more than 50 percent vote yes, we do it. If 50 percent or less, no, we don't do it.

There are some tentative results. It actually passed in two sessions. In the hypothetical sessions, 52 percent vote yes in one and 32 in another, 41 overall.

In two real experiments where you actually had to pay the 10 bucks, 24 percent said yes and 32 percent in another.

DR. PORTNEY: Two minutes.

DR. CUMMINGS: Two minutes. An agenda for further research. This is going to have to -- this is going to really be short.

One is look for new valuation institutions. I talk about this more in the paper. So, I won't spend time on it.

We don't have to feel limited. If you look at what has been going on in experimental economics over the last, even, very few years, okay, we are continually coming up with new valuation institutions that can be appropriate, you know, for different kinds of valuation situations. Okay.

We don't have to be stuck with -- we don't have to act as if we are stuck with the institutions that we now know. If we start thinking this way, all right, my feeling is

there is a great deal of optimism for developing new institutions. In the paper, I describe one that Glenn and I are playing with.

The second has to do with -- you know, it's so exciting. When you look at results like this, you don't kind of think, as is reached by some of my colleagues, it means, you know, the CVM is dead, doomed, it's not worth looking at, it's a piece of junk.

That just isn't true. Okay.

It says -- it says, "Here is where we ought to have our attention focused." All right.

And, if you look at the people that are doing work in this area and some of the results that are coming out of it, it is my view that there is just every reason for optimism. I just know there is going to be a hell of a breakthrough in two years from now.

Trust me. Okay. I will buy you a marguerite if this doesn't happen.

(Laughter.)

DR. CUMMINGS: Two years from now, all right. The state-of-the-art is going to be considerably advanced.

One thing -- one way to move in this direction is to take lessons learned from experimental economics and apply them. What's our problem?

Our problem is trying to get people on the CVM to think about opportunity costs, to report the valuation in an opportunity cost context that is somewhat akin to what you are doing when you are actually spending your money. Agreed?

Well, I said, "Why don't I tell somebody that?" I went back to, you remember, the calculator experiment. Well, before I asked the hypothetical question, I just made these four statements: All it says is to think about opportunity costs.

You know if you spend money one way, that's money you ain't got to spend somewhere else. Right. That is what opportunity cost is for.

Are you ready? Get ready, guys. Look at these

results. Whoa! Look at that.

Remember the old hypothetical in the experiment. You got rid of the wedge with cheap talk. All right.

I was so excited. And, I was a legend in my own mind for three days.

(Laughter.)

DR. CUMMINGS: That is until Lisa Rustrom at South Carolina replicated the experiment and the results didn't replicate. All right.

So, as I argue and argue in the paper, there are -- we are becoming increasingly aware of means by which we can use experiment techniques to try to shape instruments, okay, that can possibly, very possibly in my mind, remove the wedge between hypothetical response.

And, I'm done. Did I do it?

DR. PORTNEY: Terrific. Well done. Any of you who were familiar with the character, Lieutenant Frank Drevin, in the movie, "Take a Gun," know how important it is to get this portable mike back from each speaker.

(Laughter.)

DR. PORTNEY: I also want to say, before I open the floor up to questions, that I want to thank all of our presenters and our discussants. Keep in mind that each of these discussants has written a paper of his own or of her own.

And, to sort of limit somebody who has taken the time to write a paper to 15 minutes is a big burden. Everybody has really complied with that, for which I am most appreciative.

So, let me turn the floor over. And, I suppose I should first ask Jordon if he would like to respond to anything that he's heard or do you want to wait until we have taken some questions?

DR. LOUVIERE: I will wait.

DR. PORTNEY: Okay. The floor is open for questions.

Barbara.

ATTENDEE: I guess this is for anyone on the panel. I agree with you completely, Professor Louviere, that we are

never going to find one number that you can use for calibration for all these studies.

But, I think the evidence has borne that out. We've got people saying that revealed preference is greater than stated preference and such.

One of my questions is why we are doing these studies to compare the two and then just stop at the ratio and say, "Well, the ratio is this." And, why are we not going further and saying, "Why do we get this ratio?"

And, in particular, I think we need to ask what assumption we are making of the underlying response behavior when you say that that relationship is more duplicative. And, I don't think we have really thought through that.

And, then I would like to point to a very small literature that I am involved in with several other people in the room where we are trying to look at, okay, how people respond to these questions; and, what does that mean for the populace that is saying yes and no? And, can we model that directly in the likelihood function?

And, I think this might be a promising direction to take and try to explain why are some people responding to these questions differently because they have been explained the characteristics. And, I guess, just to take a central question, do any of you see this kind of thing as a promising direction?

DR. LOUVIERE: We have had a number of conversations about that very issue in the paradigm that I am associated with. As you can tell, we've really only begun to develop these results over the last four years.

We now think the most promising direction, in addition to beginning to sort out which response measures seem to be the most reliable for particular applications, it's clear there must be individual differences in these skill constants. That is, clearly different subjects are going to show different levels of variability.

That opens up a very exciting possibility of beginning to try to understand what types of individual differences account for these differences in individual liabilities. I

mean, one obvious hypothesis that has been bantered about and discussed by Professor Schumann today is the level of education.

I'm quite confident that most of the examples that I've shown you today require a reasonable amount of literacy on the part of subjects. I'm also quite confident that a large portion of our nonresponse rate can be explained by the level of education of the subjects.

And, even within the respondents who do respond, there must be differences, not only due to education but due to other strategically interesting factors. When we will be able to report some generalities that would be useful to you, I couldn't even begin to say.

It's, you know, just far too early.

DR. PORTNEY: Okay. Jonathan.

ATTENDEE: This isn't a question. But, you are free to respond to it.

I want to just mention one other institution. I like the idea of other institutions, other ways of eliciting

values.

One that hasn't been mentioned here is all the attribute utility theory as applied in the decision analysis grew out of the same contour measurement theory. And, I think it should be included.

You can respond if you like.

DR. PORTNEY: Comments?

DR. CUMMINGS: Yes.

DR. PORTNEY: Okay. Others? Okay.

ATTENDEE: I do have a question. There is a great deal of debate about the need for very large sample sizes for contingent valuation.

It's my experience in looking at the conjoint literature that large firms and researchers are quite content with a much, much smaller sample size. Could you comment on that?

DR. LOUVIERE: Yes. That's true. But, Professor Schumann's comment, as well as several other comments that were made, have a bearing on this.

And, that is, if you are actually trying to estimate with a fair degree of precision population estimates for particular target populations, whether they are U.S., Australia, some subsegments, remember that sample sizes are independent. Therefore, yes, we live with small sample sizes, around 400, for many marketing research applications.

But, if we actually want to legitimately compare differences between subsegments and the target population, then if they are in subsegments the true expected sample sizes are the same level of precision. Now, in practice, we often don't do that.

Why not? Because of resource and time constraints.

So, often what we will do is use judgment. And, that's the kindest expression I can use. We will use judgment to fix a sample size, say, 1,200 people. And, then we will attempt to do post-op comparisons.

Now, I'm the first to tell you that the power of these post-op comparisons, of course, is completely dependent on the numbers that you happen to have lined up with in the

particular groups that you are trying to make the comparisons from. And, this is not exactly a desirable situation.

So, if you are going to do something that really matters and really counts and really make comparisons between your subgroups, I think you are still back in exactly the same game that has been discussed today. But, you do get more data per respondent. Bear that in mind.

DR. PORTNEY: Questions?

ATTENDEE: Professor Louviere, would you recommend using alternative specific constants, given that there was collinearity characteristics?

DR. PORTNEY: Did you hear the question?

DR. LOUVIERE: Yes. She asked would I recommend using alternative specific constants when there is collinearity in the data set.

Well, that's a tough question, because in the Bow River study that we did, the matrix of characteristics of each of the water destinations -- and I think there were 10 to 12 of

these characteristics -- at least two thirds of them were linearly dependent on the alternative specific constants. That is, they uniquely -- a particular profile on these attributes uniquely defined a particular alternative.

Therefore, you can't separately identify alternative specific constants from the linearly dependent attribute vector that sits out there. In those cases, it's hopeless to try.

So, you might as well do one or the other. You might as well estimate nothing but alternative specific constants or nothing for attribute effects.

But, ideally you would like to do both. I mean, I don't know what the level of sophistication of this audience is between choice models.

But, these alternative specific constants pick up a number of sources of bias that lurk in choice data that you would like to clean out the utilities from. So, where possible, you would want to be able to do that.

Now, in a choice experiment, you deliberately design

the vectors of design variables to be orthogonal to one another. So that this is not an issue for design choice experiments.

They are going to allow you to estimate both alternative specific constants and attribute effects by design if you design them correctly.

DR. PORTNEY: Bill Schultze.

ATTENDEE: Well, this is sort of a double question for Professor Louviere and Ron Cummings. When I think about designing a survey, I sort of spend a lot of time trying to describe the commodity in detail. And, I think a blender is well described. But, it's also trying to focus people on subsequent goods or other things that they might do.

And, so I would just like some comments on if you might ask people, "Gee, if you took 30 dollars, might you go out to dinner with it," or there are 18 other things they might spend their money on. "How likely, or which one of those things is the first thing you would spend your money on?"

What would you suggest might be things that could be

added? And, I say that because you don't ask CV questions.

Back in 1973, we asked CV questions after two or three pages of introduction. You know, now it's getting up to 14 pages of introduction.

Anyway, could you discuss that maybe?

DR. CUMMINGS: A couple of responses. First, it's my view that probably one of the aspects of the CVM valuation or the things that we understand the least has to do with the issue that was really emphasized in the NOAA panel report but was dropped in the NOAA rule-making.

And, that is, you will recall that the panel gave particular emphasis to the need to deal with the problem of budgetary substitutes. What has never been addressed is -- or what we don't understand well is what is the range of substitutability.

If we are talking about a particular environmental good, are relevant substitutes like environmental goods, all environmental goods, all public goods? You will recall that this was addressed in Hanemann and Loomis in their JEEM

paper that appeared last year.

We have got one that follows that up, replicates their experiment, only with different goods that's in the current issue of the AJAE that you might find interesting.

Let me tell you the aspect of that that I think is -- that we need to be addressing. And, that is, both in the NOAA panel report and in the new DOI guidelines, the language that they use is that budgetary substitute should be mentioned. The question is: Is mentioning enough?

I know of no study that has really addressed this other than a very anecdotal one that is actually being -- it's under a review process, you know, right now, where one group of subjects -- you want to value Good A.

One group of subjects value Good A and seven other goods simultaneously. Another group values -- before they value Group A, the other seven goods are described to them that they might want to buy.

The question: Do you get the same response? Does mentioning budgetary substitutes give you the same result

as, you know -- I mean, hell with mentioning. You know, you are going to value them, place your value on them.

And, they find a substantial difference. I don't know if that's enough of a response for you.

The question that you raise is a very interesting one. There is one other response. But, let me just quit now so that someone else can follow that up.

Or, I will follow it up with you outside. There is another study that you might be interested in.

DR. PORTNEY: Jordon.

DR. LOUVIERE: The parallel result that we are often interested in in marketing, particularly in the last three years, to a CVM estimate is to attempt to estimate the equity of a brand. Many of you would probably call this good will.

Now, that's about as intangible an asset as I can think of trying to value. In those situations, it seems to me, and what we do in practice is to mutually, exclusively and exhaustively enumerate all possible goods and substitutes.

If not absolutely explicitly as an alternative that people can choose to purchase or not purchase, there is a catch-all category in which literally every single option is, in fact, listed. I think the experimental literature, particularly in psychology, is fairly clear that if you don't do this, particularly in allocation and related tasks, you will get things that simply don't add up.

We find, when we actually add the all-inclusive -- think of it as an all-other -- and literally name and identify explicitly everything, we reduce this absolutely dramatically. So, when things matter and count, that's typically what we do.

The example experiments you saw here, however, are of the form that the subject is explicitly told, "If these two, in general, were your only alternatives and your option was to either forego the entire experience or choose one of these, what would you do," then we don't find that to be so critical in that particular application.

DR. PORTNEY: Glenn.

ATTENDEE: I would like to get, Alan, if he could, to just expand on some of the criticisms he has in his paper and he alluded to in his talk in a limited amount of time on, perhaps if I read it correctly, the relevance of doing experiments with private goods. I obviously would disagree.

But, I'm interested, I think, just in the debate; and, particularly in terms of what Ron was saying, the paradigm experiments in economics is that if some institution has claimed to work in a complex -- there is a general claim of relevance, then it's supposed to be valid, empirical strategy.

So, I'm interested. And, perhaps you could present your criticism and give Ron a chance to discuss it.

DR. RANDALL: Sure. Well, firstly, a brief response to a comment of Ron's.

I certainly, in principle, favor experimentation. I certainly, in principle, favor hypothesis testing.

The problem arises only with respect to what interpretations are drawn when the test is over. That is,

to what does the test apply.

And, I guess coming from the business of a contingent policy referendum, which as long as I have been interested in that way of trying to get value reports, I have always assumed explicitly and in writing that it is understood by the respondent to be part of an information-gathering exercise by someone in power to decide. Okay.

The only incentive compatibility claim for a referendum is not related, Ron, to its dichotomous nature nor is it related to the fact that it's a referendum. But, it's related to the fact that the result might lead to action which might then come back effectively to raise your taxes or the prices you pay or whatever.

This condition is perfectly held in the case of the contingency policy referendum. But, I would argue, and I certainly would take some support from Rich Carson this morning, that at least there is, in many cases, a plausible, positive relationship between the voters' response and the aggregate willingness to pay and the likelihood of

implementation.

And, in those cases, I would argue that the contingent policy referendum is not purely hypothetical. And, therefore, tests about purely hypothetical things don't apply to that.

And, that's the extent of the argument. That is, the things that you are testing in the stuff that I've seen appear to me to be tests of what they are. They don't appear to me to be tests of a contingent policy referendum, as I understand it.

And, I think we will collectively, I think, make more progress when we can agree in advance on some tests to run and perhaps even commit ourselves in advance to some interpretations we might give to the matrix of possible results.

DR. CUMMINGS: I guess, as is usual, Alan, when I hear you talking, I kind of automatically say, "God damn, that's right. You know, it sounds good to me."

(Laughter.)

DR. CUMMINGS: And, again, I loved everything you said. I guess -- but, I guess I've never seen it.

You know, unless the concern that you and me and others have written about concerning the importance of a provision mechanism, unless that's not relevant anymore, how you make clear the relationship in a credible way and where, you know, you can't lie to your subjects, how you make clear the relationship of this, how this information-gathering exercise, you know, that you describe is going to work its way through, you know, to become policy and then result in a -- it's going to have be a change in the tax structure, maybe that can be done. But, I haven't seen it.

But, let me join you -- and I think we are joining Danny Connivan in what I think would just be wonderful. If we could get together, debate, set out a questionnaire design that addresses many of the issues, you know, that have been raised here today and then go out and apply it so that we are not always arguing about it after, you know, one person has taken it and done their thing and someone else

has taken it and done their thing.

DR. RANDALL: One quick sentence, if I may. I think the contingent valuation community should be eternally grateful to Ronald Reagan who has educated all Americans that policies are not free, that the bills come home to roost right on your doorstep.

I mean, the little old lady that was reputed to have said, "I never vote. It only encourages them," I think you add substantially to that. Never vote yes to a CV referendum. It sure, as heck, encourages them.

DR. PORTNEY: The last question.

ATTENDEE: It's actually a brief comment. The discrepancy between the open-ended question and the referendum result has been attributed to the difference in incentive compatibility.

It may be worth pointing out that there have been exactly the same differences in the situation where incentive compatibility is not the issue if we ask people to estimate uncertain quantities under conditions where the

error is probably not normally distributed. So that when you ask the length of the Mississippi or what is the height of the tallest redwoods, or questions of that type, you can obtain the distribution of people's beliefs about those quantities in either one of two ways, by asking them to produce an open-ended estimate or by asking a series of referendum questions and reconstructing the distribution of estimates from that.

When you perform the two experiments -- and I have done that -- you then obtain results that are undistinguishable for willingness to pay and for information. The same upward bias that has been pointed out in the context of willingness to pay appears when you are estimating the length of a river.

And, it's entirely useless, as far as we can see. The point is worth noting, because in the question of information, you will never hesitate. It would be obvious that if you want to know people's beliefs about an uncertain quantity, you just ask them the open-ended question.

DR. PORTNEY: Responses here?

DR. RANDALL: I don't know for sure about that. There seems to be a lot of different stories getting around that we need to get to the bottom of in terms of the dichotomous referendum versus open-ended, whether it goes higher or whether it goes lower, as being nay-saying stories offered here today.

In Ohio, pretty much everything financial, it seems like, is decided by referenda. It's impossible to raise school or property tax collections even as a result of increased assessments without putting it to a vote.

After assessments, the tax rates are calculated down. And, if you want another opinion, you go to a vote.

It's widely believed that that nay-saying dominates, at least, around the median. Not only do we vote on everything in Ohio, but if you take the petitions around and get it on the ballot, the Electoral Commission decides the wording.

And, in one famous case, all of the interests agreed they wanted the issue to pass. And, so they decided, and

the Electoral Commission agreed, that the referendum would state, "You do not want this, do you?"

(Laughter.)

DR. RANDALL: And, if you voted no, as the majority did, it, in fact, happened.

(Laughter.)

DR. RANDALL: So, there seems to be a lot of evidence for nay-saying, at least, around the median. So, I am not sure.

There are different messages about what we are getting, particularly from the referendum, depending on what the experiment is or what the real natural experiment is.

DR. PORTNEY: Jordon.

DR. LOUVIERE: I wanted to make two unrelated comments to this. One is that I've been struck in reading your literature that no one seems to pay much attention to warm-ups or practice.

Let me scare you greatly by telling you what a former Ph.D. student of mine, Robert Meyer, who is now at Wharton

did, in the late 70s. It's published in a place that remains totally obscure but needs to be read by lots of people.

What Bob did was to vary bundles of goods in different order conditions using a Latin square where the bundles were constructed using these conjoint type experiments so that he could actually look at the effect of each attribute at each order in which a profile was shown. Now, under the hypothesis that the first time you show somebody something their utilities or preferences are stable, you would expect that the effects of these attributes should not vary except from statistical sampling error regardless of when you show them to them.

If you take all of the profiles or bundles that people saw in order one and you estimate a model and you take all the ones in order two and you estimate a model, you should get the same utilities. That's not what happens.

In fact, what happens is the effects bounce just like this across all of the attributes for the first five to six

bundles that people see. And, then they are remarkably stable thereafter.

We have used this finding for years. In those choice sets that I showed you, we typically give people four or five up front that we throw away. We don't even analyze them, because we know that people use them to learn.

I find it inconceivable that none of you seem to have picked up on the fact that if you give people something once, especially if it's highly complicated, they are going to do something totally different than if you give them something that's even slightly different the second time, because they are going to learn something different when you give them the new information which may cause them to reevaluate their preferences. So, one reason you are going to get variance in results is simply by virtue of the fact that you are not giving people practice in the actual task that you want them to learn.

In addition, there are all sorts of other effects that occur that we try to counterbalance for in these kinds of

experiments that we run. It's not only practice but, you know, particularly for those of you that like to publish aggregate choice data, in my area we worried one hell of a lot about a phenomena in aggregate choice data called heterogeneity preferences.

If you've got two sub-samples in your population with totally different tastes in preferences, that is different weights on the utilities for different attributes, and you give them, first, one profile and you measure the aggregate choice distribution and you report that in a paper, I can assure you that if you give them a second profile that aggregate choice distribution changes if for no other reason than you change basically the preference distribution in your population. So, that now, let's say, that 40 percent had one value system and 60 percent had the other.

In the first result that you publish, as soon as you give them the next profile, which might favor the group with 40, you get a totally different result which could look like it violates regularity and all kinds of things that have

been reported in the literature and suggested in aggregate data or violations of economic assumptions or violations of transitivity and all sorts of other things. They are probably just violations of the fact that you didn't take into account the fact that individuals are different.

We worry about that a lot. None of you seem to worry about that very much except that you often incorporate individual difference terms into your models, things like income and age.

But, I hate to tell you, that's observed heterogeneity. You have to care about unobserved heterogeneity. And, I don't see that being captured in any of the models I've seen published in your literature.

Let me issue another invitation. Every three years, we put on a joint conference with a number of psychologists, a number of very, very top econometricians in the choice modeling field, marketers in transportation people.

Richard Carson came to this conference that we held at Duke this summer. And, we benefitted an enormous amount

from Richard being in our workshop for the four days that he was in there. And, I think it's fair to say, right, Richard, that Richard also benefitted a lot from listening to our perspectives.

And, I think a number of the people in this room that are doing research like this, which is very compatible and similar to the research that is now being done in economics, psychology and marketing would benefit from coming to this conference. If you are interested, the next one will be in three years.

Please, drop me a line. I will put you on the list.

DR. PORTNEY: Jordon, thank you very much. I want to call today's proceedings to a halt.

Dinner will be in 45 minutes in this room. And, the sooner you get out of here, the faster they can set up for dinner.

Please, join me in thanking our three final panelists here today.

(Applause.)

DR. PORTNEY: Thank you.

(Whereupon, the meeting is recessed at 5:50 p.m., to reconvene at 7:53 p.m., this same date, for the dinner speaker's presentation.)

EVENING SESSION

(7:53 p.m.)

DR. BJORNSTAD: May I have your attention, please? Thank you.

It's my pleasure to introduce our speaker for this evening. And, before I do that, please let me introduce his guest, Senator Linda Hinsland of South Dakota.

(Applause.)

DR. BJORNSTAD: We are very happy she could join us. Dr. Thomas E. Lovejoy is the Assistant Secretary for Environmental and External Affairs at the Smithsonian Institution.

By training, he is a tropical biologist and a conservation biologist. But, his work takes him well beyond that.

He has done such things as introduced banding of birds in Brazil, introduced the concept of the minimum critical size of ecosystems, conceived the notion of the trading depth for nature. And, he is the founder of the Public

Television series, "Nature."

In addition, prior to being at the Smithsonian Institution, he was with the World Wildlife Fund. Currently, he is advising several agencies of government, including the U.S. Secretary of Interior.

He is the past President of the American Institute of Biological Sciences, the past Chairman of the Biosphere Program. And, the list goes on and on.

So, please, without any further ado, let me introduce Dr. Thomas Lovejoy, our speaker for this evening.

(Applause.)

DR. LOVEJOY: Well, thank you very much for this somewhat daunting opportunity to show how little I actually know about economics, but to talk about economics and biology and perhaps give some examples that will give you food for thought as you consider the way economics and environment interact in constructive ways and sometimes not so constructive ways.

First of all, a word or two about sustainable

development. I have colleagues who believe that sustainable development is oxymoron, that it is absolutely impossible to have sustainable development, and that economic growth is just inherently destructive of the environment.

I really disagree with that. However, we see plenty of evidence of the destructive ways that economic activities can have.

And, one of the reasons I disagree with it is that I feel that there has to be in economic growth parallels to the two kinds of growth you have in biological systems. One type of growth is where the organism simply gets bigger and bigger, the kind of problem I am always struggling with.

(Laughter.)

DR. LOVEJOY: And, that's really analogous to economic activity that requires an increasing amount of natural resources to grow.

The other kind of biological growth is known by the rather esoteric word of "growth by intussusception." And, what that means is growth in which the organism becomes no

larger but it becomes more complicated and more complex.

And, it's sort of growths or parallels in economic growths. It has to be the kinds of activities that really require relatively little in the way of natural resources -- activities like service industries or a wealth of possibilities in the information industry.

I also would say that as we sort of try and understand what sustainable development is that in the end an important piece of it is going to be biological in nature, simply because of the inherent tendency of all living organisms to make more of themselves. In fact, in referring to that kind of capacity of renewal in environmental resources, I got one of Greenwier's Quotes of the Year in 1993, which was a simple little quote, "Thank God for sex."

But, I'm quite serious about the role that biological resources have to play in sustainable development. And, one of the examples I plan to give you tonight will talk about how that is now beginning to happen at extremely sophisticated and complex levels.

What I wanted to do was, first of all, talk about some examples, real examples from the real world, in which there are different ways in which economics could look at the situation. And, there are so many things about economics which really are not particularly disposed in the way economics normally is used to treat the environment properly, such as all the long-term values of biological resources which just go against the grain of the way discount rates are used.

But, there are also very often -- and I'm sure you will find it no surprise -- a whole series of different ways economics can look at the same resource. We are not very far here from the greatest estuary of the North American Continent, namely the Chesapeake Bay.

And, there is a whole variety of species of plants and animals that live in the Chesapeake Bay. The watershed is huge.

It goes all the way up into New York State, encompasses a number of states. And, one species in the Chesapeake Bay,

for which it is duly famous, is the edible oyster.

One way of valuing that species is as a single, potentially sustainable resource in terms of seafood. That, you know, is one way you can put a value on oysters and the production of oysters in the Chesapeake Bay.

A completely different measure one could give to the value of oysters represents the oyster as an indicator species of environmental change. And, I'm thinking, in this case, of the antifouling compound to tributyltin which used to be put in nautical paint so that barnacles and other things of that sort would not grow on the bottoms of boats belonging to private boat owners or the Navy or whatever.

And, it turns out that the first indication that tributyltin or TBT, as it's known for short, was not necessarily such a good idea to have seeping into the watersheds of this country was deformed aquatic organisms. The first sign of it actually showed up not in Chesapeake Bay but in oysters off the coast of France.

In Chesapeake Bay, oysters were not the first to show a

problem. It was another mollusk, a whelk, which basically was affected in its reproductive system by this compound, such that female whelks grew penises in a condition known as impo-sex.

This is one of the very first of the sex-altering compounds to be -- seen to be out there in nature affecting things. In any case, the value of the oyster as some kind of indicator of that kind of environmental change is yet another way one could put some number along side the oyster.

But, there is yet another. And, it involves the filtering capacity of oysters, of that single species.

Today, the oyster population of Chesapeake Bay filters a volume of water equal to the entire bay about once a year. Before the water chemistry of Chesapeake Bay was severely altered, primarily still by agricultural runoff, before that, the population of oysters in Chesapeake Bay filtered a volume equal to the entire bay once a week, having dramatic impacts on water quality, on the way the whole Chesapeake Bay ecosystem and the source of so many marine resources

actually function.

So, the point that I am trying to make here is there are three different ways you can look at putting a value on the oysters in Chesapeake Bay. One is a food source. One is sort of an indicator of environmental problems. And, the third as a major component in ecosystem process.

And, of course, nothing in the way that economics has worked in the Chesapeake Bay region has put any value at all on the oyster other than the food supply approach.

A different kind of example would involve the floodplains in the Amazon Basin, which is an area of particular interest to me. And, the floodplains in the Amazon Basin are attractive because, amongst other things, they are one of the few places you can decent soil for agriculture because of the annual deposit of silt.

They also support floodplain forests, are suitable for extraction of timber from the forest with relatively less disturbance than timbering a dry land Amazonian forest. But, far more important than that, in many ways is their

link the Amazon fishery and one of the most interesting bits of natural history that I know of which relates to economic value.

The Amazon is one of the wettest places on earth, the Amazon Basin. The river system holds 20 percent of the world's river water.

And, during the rainy season, so much water falls that the level of the river rises something like 10 meters or more, a good 30 to 40 feet. And, all those floodplains and their forests get flooded.

And, what happens is that many of the fish species then swim into the flooded forests and feed on foods and seeds and other organic matter which falls into the water. And, basically, for many of them, that's their only food supply for the entire year.

I mean, even piranhas go vegetarian during the high water months of the year. And, something like three-quarters of the commercial fish species of the Amazon depend on this rather curious link between the forests of the

floodplains and fishery production.

So, again, there are various ways you could look at the value of the floodplain. You can look at it in terms of potential agriculture. You can look at it in terms of extraction of timber or as the ecological base for much of the Amazon fishery.

And, then you have to sit there and try and sort out those different values into a coherent land use policy.

But, the next example I would like to give you is how a single resource can continually be amplified in its importance to human society. And, how all of you could deal with that is beyond me.

But, in any case, in 1993, the Nobel prize for chemistry was shared by a Californian named Cary Mullis, who got it for conceiving of a reaction in molecular biology known as the polymerase chain reaction or PCR for short. That is not a household phrase or acronym.

But, it has a great deal more to do with daily life in the United States these days than most people have any clue.

This is a reaction that allows the genetic material of whatever it is that you are interested in to be multiplied over billions and billions of times in the space of just an hour or two.

It is the only true part of the science in Jurassic Park. It is the science which makes possible esoteric things like really identifying the last remains of the last of the czars of Russia.

But, it is also central to modern diagnostic medicine. And, you no longer have to wait for that throat swab to be cultured two or three days to get the identification of the offending organism and, therefore, the prescription and get yourself back on the road to health and being productive in the economy.

This is a reaction which is central to the whole field of biotechnology, much of the research of molecular biology. And, thinking of it just in those terms alone, this reaction is probably generating something on the order of five to 10 billion dollars of economic activity every year.

More recently, I've learned this week in a wonderful presentation on the human genome project that the entire project to characterize our genetic structure is based squarely on this polymerase chain reaction. So, all that ambitious project means, in terms of advances in human health, which are extraordinary, depends again on this abstruse reaction.

The reaction works because of two things. One is the application of heat, which causes the two strands of the chromosome to unravel; and, then an enzyme, which causes each strand to create the missing strand.

So, you go from one to two and then from two to four and four to eight, et cetera. And, Cary Mullis dreamed up this reaction. But, nobody had an enzyme which would do the replication part which would not be sort of denatured by the heat part of the reaction.

So, you had the reaction. But, it wasn't a chain reaction.

And, ultimately, somebody had the bright idea that

surely somewhere in nature there are organisms living at extremely high temperatures which might well have an enzyme, a heat resistant enzyme, that could do the job. And, so somebody went out to Yellowstone National Park, looked in the hot springs, looked in the slime of the hot springs, which I like to say any decent mother would have taken Ajax to, and there they found this little bacterium which is called thermocyclaticus (phonetic) and it has the molecule which resists heat and does the replication job and makes this whole polymerase chain reaction and all that it means for human society and economics work.

And, it is sheer luck that Yellowstone National Park happened to be set aside as a protected piece of landscape because it was attractive in 1872. There was no, sort of, vision at the time that biological diversity might include species like this little bacterium, which nobody could have seen in those days anyway, that would have the power to transform society.

So, I think one of the challenges that we need to think

about in biology and economics is just how you approach this problem where a single organism or a single molecule out of nature makes something possible that a whole series of other things stem from and continually amplify in terms of economic activity and benefit.

A couple of more points. One is, the whole area of the functioning of ecosystems and how the species of plants and animals in the ecosystem relate to that functioning. The big question in science is do you need all those species to make these ecosystems function in terms of cycling water and energy and nutrients.

And, there are a lot of species in these ecosystems. Surely, some of them are redundant, or so it would seem, which is sort of edging up to the whole notion of substitutability in economics which usually gives the big shivers to biologists.

(Laughter.)

DR. LOVEJOY: The truth is that we have very little understanding about how biological diversity, indeed,

relates to ecosystem functioning. Many of these ecosystems have lots of rare species in them which just a priori you would take one look at the situation and say, "They can't be very important."

But, you have to remember that the kinds of things that live in an ecosystem are not only reflected by what the ecosystem is doing at the moment you are looking at it but also are a reflection, one, of its history and, two, of fluctuating conditions within which that ecosystem exists over time. And, I will give you a really wonderful example that came out of a laboratory of the wonderful woman limnologist, Ruth Patrick, now age 86 and about to publish five volumes on the rivers of the United States and their biological diversity, an extraordinary lady.

Somebody in her laboratory studied a tiny little fresh water yeast, a really weirdo organism which didn't have a normal metabolic pathway. Normally, it was very rare in these aquatic ecosystems.

And, what was abnormal about its metabolic pathway was

that it skipped certain standard steps, steps which are vulnerable to mercury compounds, mercury poisoning. And, what happens with this little yeast when there is natural or unnatural mercury pollution in aquatic ecosystems is all the other species suddenly are at a disadvantage. And, this one is sort of king of the roost.

And, it explodes in population. It reduces those mercury compounds to quicksilver, builds up a store of it in its vacuole and then eventually sort of sticks a little dab of quicksilver on a nearby rock and ultimately cleans up the ecosystem of mercury, puts itself at a disadvantage and becomes rare again.

So, one of the things we have to be really careful about is thinking about ecosystems as though they only exist in the state we are looking at them at the particular moment. And, I leave it to you in sort of nightmares to figure out how to put value on all of that.

(Laughter.)

DR. LOVEJOY: People have talked for years about

ecological systems and the more diverse ones having more stability. They have talked about them having more resiliency.

And, basically, I've got to tell you, there was zero evidence to support any of that until this last year when some work done by David Tillman and colleagues at the University of Minnesota, looking at a series of vegetated areas and their response to drought years, found that the ecosystem with the highest level of biological diversity, the closest to the natural system, was the one that weathered -- if I may use that terrible pun -- that ecological stress most easily.

And, so I just mention that as one of the other important things we need to think about.

Lastly, there is the following: If I am right, which I really think I am, about sustainable development depending, to an important degree, on biological resources, then the life sciences and all they can mean for us become extraordinarily important, more important than they ever

have been before.

And, in the last analysis, the life sciences advance by studying how biology works in the whole variety of life on earth, including little weirdo yeasts which like to eat mercury. And, we are still learning some of the most fundamental dimensions of life on earth.

I mean, 15 years ago, we did not know about the biological communities which live near the thermal rifts in the bottom of the ocean and basically depend not on sunlight but on the primal energy of the earth. And, we are still discovering major new forms of life.

I mean, it's different from ourselves as we are from earthworms, for example. And, the point I'm really leading up to is that the variety of life on earth is, in the end, a fundamental library on which the life sciences are built.

And, it makes no sense for us to be allowing many species to just disappear to extinction when there is so much that we can learn from them and benefit from them. How do you value that?

Well, I suppose there is something to be learned from how we value libraries. I'm not talking about valuing a book because it's a rare book. I'm talking about how you value books because there is information in them.

Anyway, that is just a series of examples that I thought would allow you to be a little distracted from contingent valuation and think about how economics and biology intersect.

Thank you very much.

(Applause.)

DR. BJORNSTAD: Does anybody have any questions before we close that they would like to ask Dr. Lovejoy?

(No response.)

DR. BJORNSTAD: Thank you very much. Then, we will close for the evening.

(Whereupon, the meeting was adjourned at 8:20 p.m., Thursday, May 19, 1994.)