UNIVERSITY OF WAIKATO

Hamilton
New Zealand

THE REFERENDUM INCENTIVE COMPATIBILITY HYPOTHESIS: SOME NEW RESULTS USING INFORMATION MESSAGES

Gianluca Stefani

Riccardo Scarpa

Department of Economics

Working Paper in Economics 10/07

June 2007

Gianluca Stefani
Agricultural and Resource
Economics Department
University of Florence
P.le delle cascine 18
50144, Firenze, Italy
Tel: +39 (0) 55-3288-223
Fax: +39 (0) 55-361771
gianluca.stefani@unifi.it
Web: http://www.deart.unifi.it

Riccardo Scarpa
Economics Department
University of Waikato
Private Bag 3105
Hamilton, New Zealand

Tel: +64 (0) 7-838-4045
Fax: +64 (0) 7-838-4331
rscarpa@waikato.ac.nz
Web: http://www.mngt.waikato.ac.nz

Abstract

We report results from a laboratory experiment that allows us to test the incentive compatibility hypothesis of hypothetical referenda used in CV studies through the public or private provision of information messages. One of the main methodological issues about hypothetical markets regards whether people behave differently when bidding for a public good through casting a ballot vote than when they are privately purchasing an equivalent good. This study tried to address the core of this issue by using a good that can be traded both as private and public: information messages. This allows the elimination of confounding effects associated with the specific good employed. In our case information dispels some of the uncertainty about a potential gain from a gamble. So, the approximate value of the message can be inferred once the individual measure of risk aversion is known. Decision tasks are then framed in a systematic manner according to the hypothetical *vs* real nature of the decision and the public *vs* private nature of the message. A sample of 536 university students across three countries (I, UK and NZ) participated into this lab experiment. The chosen countries reflect diversity in exposure to the practice of advisory (NZ) and abrogative (Italy) referenda, with the UK not having any exposure to it. Under private provision the results show that the fraction of participants unwilling to buy information is slightly higher in the real treatment than in the hypothetical one. Under public provision, instead, there is no statistical difference between real and hypothetical settings, confirming in part the finding of previous researchers. A verbal protocol analysis of the thought processes during choice highlights that public provision of information systematically triggers concerns and motivations different from those arising under the private provision setting. These findings suggest that the incentive compatibility of public referenda is likely to rely more on affective and psychological factors than on the strategic behaviour assumptions theorised by economists.

Key words:

Contingent Valuation

Incentive Compatibility

Experimental Economics

JEL Classification

Q50, H40, C91

1 Introduction

Contingent valuation methods are widely used in nonmarket valuation. However, CV has been criticised for a number of reasons revolving around the issues of credibility and reliability of responses. In particular, the potential incentive compatibility of hypothetical referendum format is a key element of validity for nonmarket value estimates obtained from CV surveys. The aim of this research is to investigate the impact of the private vs public nature of the good on the incentive compatibility of hypothetical referenda for obtaining estimates of monetary values for nonmarket goods and services. One novel aspect of this study is the use of a good which is neutral with respect to the public or private provision mechanism. This allows us to compare public versus private provision rules and their impact on incentive compatibility without suffering the confounding effect on the outcome associated with the use of an eminently private good. Such effects have previously detracted from otherwise high profile studies. The provision-neutral good used here are information messages able to resolve part of the uncertainty of a gamble. Results are of interest for applied economists and practitioners working for government agencies that use cost-benefit analysis of environmental, health or other public programs.
The paper is set out as follows. In the first section the research questions and the objectives are presented along with a review of the literature and a synthesis of the findings to date. The methodology and data collection are illustrated in the second section. The analysis of quantitative data is summarised in the third section while a qualitative analysis of a verbal protocol from respondents is illustrated in the fourth section. Conclusions are drawn in the last section, along with suggestions for further research.

2 Objectives

Contingent Valuation (CV) is a survey based method employed to elicit respondents' willingness to pay for unpriced goods contingent to the existence of hypothetical markets. More than two thousand papers have appeared on the topic and both government agencies and international organisations have promoted CV studies across over 50 countries (Carson 2000). Despite its wide success as a preference elicitation method CV has been criticised for a number of reasons revolving around the issues of the credibility and the reliability of responses (Cummings et al. 1997, Diamond & Hausman 1993, Green et al. 1998, to name but a few).
One of the possible drawbacks of the method refers to the incentive compatibility of the format used to elicit preferences. Carson & Grove. (2007) showed that only the referendum choice format with its binary discrete choices (such as the referendum format) can be (but not necessarily are) incentive compatible. In particular, hypothetical surveys can be incentive compatible provided that: 1) respondents believe that their answer will affect the actions undertaken by the agency in charge of the provision of the good (that is the survey is "consequential"), 2) a "one shot" issue voting procedure is adopted. Thus, while for public goods it is possible to make people believe that the provision of good will be influenced by the advisory referendum results in other cases such as in the provision of new private goods, payments cannot credibly be enforced and hypothetical questions related to willingness to pay (WTP) are not incentive compatible.

However, other authors (Cummings et al. 1995, 1997, Harrison & Rutstrom 2002) point out that researchers cannot know in advance whether a respondent to a hypothetical survey feels that the outcome of the survey will affect her utility through the provision of the good, hence, the consequential nature of the survey must be assumed for public goods too.
Results from the literature that investigates this issue empirically seem unpersuasive due to the failure to address the exact research question provided by the theory on incentive compatibility. Thus the hypothesis remains, so far, largely inadequately tested.
The incentive compatibility argument in support of CV hypothetical referenda provides researchers with several testable hypotheses that, so far, have been investigated in settings where confounding factors (such as the nature of the goods for which preferences are elicited) often blurred results. One testable hypothesis states the lack of incentive compatibility of hypothetical referenda on the provision of new private goods vs. the presence of incentive compatibility for public goods. Results on tests of this hypothesis would be more persuasive if they were performed using a good that could credibly be purchased privately and collectively, while maintaining the same private value to individuals. Information messages have this desirable property (Hirshleifer & Riley 1992). A message can be acquired privately, but becomes a public good as soon as its content is revealed to other people. Moreover, information maintains the same private value for each subject under different provision rules as it is valuable only indirectly. The ex ante value of information messages is given by the expected gain from making informed vs. uninformed choices in a game against nature. Furthermore, the value of information may be characterised, at least partially, as a 'homegrown' value (Cummings et al. 1995) since in a risky environment, and for a given degree of risk induced by researchers, the value depends on the attitudes of respondents towards risk (Lawrence 1999). This is, for instance, what is observed in markets for insurance (see for example McClelland et al. 1993, for some relevant results). However, attitude towards risk is related to the functional form of the elementary utility function. Individual utility functions can be investigated by making subjects performing a series of choices or stating certainty equivalents in a series of gambles (Becker et al. 1964, Davis & Holt 1993, Currim & Sarin 1989, Harrison 1986, Hey & Orne 1994, Camerer 1995). Assume that the curvature properties of the utility function are sufficiently known, than it becomes possible to derive the value of information in a specific game setting. More generally, measures of risk aversion of subjects are expected to be correlated to the values subjects pose on information messages capable to mitigate, at least in part, the uncertainties surrounding the choice task they are facing. This may suggest a further test for the demand-revealing properties of the referendum mechanism.
In this study we make use of these concepts and in our study the public provision of information is attained within a controlled environment as information messages are of no use to participants outside the context of the experiment. Hence, we do not expect to observe free riding in the experimental referendum setting.
The aim of this research is to investigate the impact of the provision mechanism of the good (private vs. public) on the incentive compatibility of hypothetical referenda while controlling for the perceived consequentiality of the survey. The focus is on eliciting values for a good, such as information, that has a double public-private nature, thus allowing a way to compare public versus private provision rules and their impact on incentive compatibility. A secondary goal is that of using verbal protocol

analysis of thought lists to identify thought processes systematically associated with the two decision settings.

3 Research design

The nature of the objective suggests the adoption of a conclusive research design. A causal study was designed and undertaken to test the effects of hypothetical vs. real choices under private and public provision on the decision to purchase information messages.
Previous testing of the presence of hypothetical bias has largely relied upon valuation experiments. Experiments allow a better control of the environmental variables enabling researchers to concentrate on the impact of the treatment variable. In this case these are the hypothetical vs. real nature of the referendum and the private vs the public form of provision.
We performed a series of valuation experiments drawing upon the protocol used by Cummings et al. (1997) as their objectives were similar to ours. However, the nature of the good provided required to be changed. In particular, the underlying choice setting was a simple two-state of the world, two-message, two-action decision problem, similar to that employed by Wendt (1969).
It is to be emphasised that the hypothetical nature pertains only to the buying or selling of information, but not to the outcome of the game to be played, which involved participants playing a game for real returns in a one shot context. This allowed us to enforce consequentiality with respect to the decision whether to buy or not the information message.

Subjects recruitment

Participants were recruited mainly among undergraduate students of the University of York in UK, the University of Firenze in Italy and the University of Waikato at Hamilton in New Zealand, either by on-site recruiters or by email and web-advertisement. Five hundred and thirty six students were recruited in this manner almost evenly split across the three Universities. Nine of them did not complete all the experimental tasks and were subsequently discarded from the analysis. Of the remaining 527, 94 resulted postgraduate or older than 23. Indeed, especially in Italy, students on prolongation of their studies are not uncommon and they were difficult to identify before they completed their questionnaire. In the econometric analysis of data we accounted for this characteristic by means of a dummy variable.
Each participant was paid 10 pounds, 10 euro or 10 NZ dollars as participation fee depending on the location and was told that she (or he) could have had the opportunity of doubling the sum in the process of taking part in the experiment.
The allocation to experimental treatments was random, except for balancing between sexes. However, different showup rates sometime resulted in imperfectly balanced groups across experimental sessions. The experiment took place in university classrooms. All the UK and NZ sessions were located at central campus whereas the Italian sessions were split between classrooms in the life science faculty and classrooms in the teaching centre for the faculties of medicine, mathematics and engineering.
Following the experimental protocol employed by Cummings et al. (1997) participants took part in a market experiments prior to their participation in the core experiments. The markets experiments were simple oral double auction whereby units of a generic good were traded following the procedure set out by Davis & Holt (1993). The auction

served a double purpose. Firstly, it was meant to provide participants with some experience in participating in experiments without biasing their response in the core section of the experiment. Secondly, as the oral auction takes about 30 minutes to be completed, this made participants feel that the participation incentive they were paid was worth the time they had to spend attending the experiment.
After completing the double oral auctions, participants were provided with the dichotomous choice exercise on the value of information described in the next subsection. The subsequent two phases of the experiments aimed at eliciting a measure of risk-aversion and at collecting demographic data through a self compiled questionnaire. Risk aversion was elicited through a series of 10 choices between paired lotteries following the protocol published by Holt & Laury (2002). Each pair of lotteries was composed of a risky (B) and a less risky (A) option. The risky option initially provided a payoff of 11.55 pounds (or euro or NZ dollars) with probability one tenth and a payoff of 30 pence (or cents) with probability nine tenth. The less risky option provided respectively payoffs of 6 and 4.80 pounds (or euro or NZ dollars) with the same odd. Only a risk lover would choose the B lottery in the first decision. However, along the series of choices the probability of the high payoff in both lotteries increases (see table 1) so that subjects were likely to switch from the less risky to the risky lottery at some point. The later the switch occurs the more risk-averse the subject. Thus the switching point provides a measure of risk aversion, while the occurrence of more than one switch point signals a low level of understanding of the rules of the game or, possibly a somewhat “irrational” behaviour.

The design of the experiment

This part of the experiment aims at investigating whether participants faced with the choice to buy a piece of information behave differently depending on whether their choice is real or hypothetical, and on whether the information is purchased collectively via a referendum decision rule (provided publicly) or purchased privately. Therefore the experimental treatments were based on the combination of the levels of the following variables:

1. rule of provision of information messages (2 levels: private, public),
2. form of payment (2 levels: real vs. hypothetical).

A 4-cell full factorial design was obtained. For each cell 3 experimental sessions were replicated at each location, for a total of 24 sessions.
The underling game for which the information is valuable was framed as a crop choice problem. In the game two crops, A and B, give two different returns, depending on the prevailing weather conditions, or states (dry or wet). Participants were told that although the prevailing state of the world was uncertain, they were given the probability—or likelihood—of the occurrence of each state. In addition, they had available for a fee a weather forecast which could correctly predict the weather condition 75% of the times.
Thus subjects had to choose between two actions (growing Crop A or Crop B) the payoff of which depended on the realization of one out of two possible states of the world (wet or dry weather). The payoff matrix was chosen to be symmetric so that π(Crop A, wet)= π(Crop B, dry) = 10 (pounds or euro) and π(Crop A, dry) = π(Crop B, wet) = 0.
Subjects were then asked to participate in a training trial of the crop choice game in which they did not receive any real payment. Prior to participating in the crop choice

linked to real rewards, subjects were asked whether they would like to pay a given amount for the provision (private or public) of the information service consisting in one out of two possible weather forecast messages ("the weather will be dry", "the weather will be wet"). The likelihood of the messages given the state of the world are symmetric with Pr(weather is dry | dry)= Pr(weather is wet | wet) = 0.75 while prior marginal probabilities (Pr(dry) , Pr(wet)) are equal to 0.5. The price of the message was at 2.75 pounds (or euro). This value is 10% higher than ex ante value of the message for a risk neutral subject (2.50 pounds or euro) but would be accepted by risk averse subjects thereby increasing the role of home-grown values in determining the value of information.
The state of the weather was determined by blindly drawing a ball from a box containing an even number of black (wet) and white (dry) balls and placing it in a black bag for checking upon request later on. Similarly, whether the message was correct or incorrect was determined by drawing a ball from a cage containing 3/4 of yellow balls (correct message) and 1/4 of purple balls (incorrect message). Participants were shown table 2 to illustrate the likelihood of the message given the state of the weather.
After the information phase the attention of the subjects would be moved to the crop decision game that always implied real payments and eventually participated in the lottery game and the questionnaire phases described above.

The proposition

The information service was provided as either a private or a public good. In the public provision case participants were to vote and in case of a majority wanting the piece of information, a tax was to be collected from all participants in the session and the information was made public. In the private setting those who chose to buy the information service received it in a sealed envelope and were asked not to divulge its content. Following the protocol used by Cummings et al. (1997), participants paid real money for the information service in the real treatment while in the hypothetical one there was no such payment.
The dichotomous choice question was framed differently depending on the treatment. In the private provision case participants in the real setting were provided with the following question:

We would now like to offer you the opportunity to actually buy the weather forecast service at a price of £ 2.75. If you say "yes" then you will have to pay the amount right now and you will receive the message in a sealed envelope. You will then select the crop which you think is the best (the more profitable) based on the weather information. Are you willing to pay £ 2.75 for the weather forecast service?
Please circle "yes" or "no" when you have made up your mind.
YES NO

In the hypothetical setting the question was the same but for (a) the use of the conditional tense whenever appropriate (e.g. "you *would* have to pay", "you *would* receive", "*would* you be willing to pay") and (b) the first statement that was replaced by:
"We are not actually offering you [. . .]. However, we are very interested in your response to the following question. . . ".
In the public provision treatment participants were asked to vote on the following proposition:

On the proposition: Everyone here in the room will contribute £ 2.75 to the experimenter. The contribution is to be used for making available to everyone the weather forecast service.
My vote is

YES NO

Please circle as appropriate when you have made up your mind

In the real setting the proposition was introduced by the following passage:

"We would now like to offer you the opportunity to actually obtain the weather forecast service. We are going to have a secret vote to decide whether or not we will do this: all of us pay £ 2.75 for this purpose:
Proposition
Everyone here in the room will contribute £ 2.75 to the experimenter.
The contribution is to be used for making available to everyone the weather forecast service.
If more than 50% of you vote "YES" on this proposition all of you will pay £ 2.75. The experimenter will collect £ 2.75 from each of you soon after the vote and then the forecast message will be written on the board for all of you to read.
If 50% or fewer of you vote "YES" on this proposition no one would pay £ 2.75 to the experimenter and no forecast will be available to any of you."

Even in this case the wording in the hypothetical setting was similar, except for the use of the conditional tense where appropriate, and the replacement of the first paragraph by the following one:
"We are not actually offering you the opportunity to obtain the weather forecast service. We want you to suppose that we were to have a secret vote to decide whether or not we should do this: all of us pay £ 2.75 for this purpose. Supposing that we were to have such a referendum we would vote on the following proposition:"
Participants in the hypothetical treatments were provided separately with both the public and the private provision questions to check for consistency in answering within a given design. Of course, this was not possible in the real setting.

The verbal protocol analysis

A verbal protocol analysis was employed to understand (a) how subjects made sense of the questions about the provision of information and (b) whether there were any salient and systematic differences across the four treatments. Protocol approaches have been used in marketing (Jaeger, 2005) as well as in environmental economics (Schkade and Payne, 1994; Svedsater, 2005). In the written version of protocol analysis adopted in this study subjects were asked to write down their thoughts about the choices they had just made. This thought list technique is "a type of cognitive assessment that is particularly useful when one either has no predetermined ideas about the cognitive dimensions that are relevant, or has only a few untested hunches" (Cacioppo et al. 1999). It is worth noticing that factors of success in the thought listing protocol are: 1) answers must come soon after the choice, 2) instruction must be clear, 3) participants must respond honestly. That is why participants were provided with the form immediately after the information purchase/vote decision and before making the final crop decision. In addition, clear instructions were provided to increase the precision of answers (for the public provision session the statement was modified accordingly):

"Just before moving on to choose the final crop, we are interested in everything that was going through your mind when you were deciding to accept or refuse the hypothetical private purchase (in a sealed envelop) of the weather information. Please list these thoughts, whether they are about yourself, the situation and/or others; whether they are positive, neutral and/or negative:"
Since anonymity was assured to participants we assume that they have responded honestly. The main themes were then identified by means of content analysis on the transcripts of the box content. The coding scheme was partially built taking into account previous works on how people make sense of CV questions (Schkade and Payne, 1994; Svedsater, 2003). However, the tasks people had to perform in this experiment were quite different from those encountered in a CV exercise and an inductive approach had to be followed for most of the items. The NVivo® qualitative data analysis software was used for this purpose.

4 Data Analysis

The following null hypotheses were retained for the analysis:
H1: Hypothetical <u>public</u> provision of a good is incentive compatible hence subject behaviour, as measured by "yes" responses, is independent of the real or hypothetical referendum design.
H2: Hypothetical <u>private</u> provision of a good is incentive compatible hence subject behaviour, as measured by "yes" responses, is independent of the real or hypothetical nature of the dichotomous choices about purchasing information with private value.
If both hypotheses cannot be rejected, then this would mean that no difference exists in incentive compatibility of hypothetical settings between public and private provision formats. If only one hypothesis is rejected this would suggests that the rule of provision plays a non-negligible role in determining the incentive compatibility of the instrument.
Experimental data are reported in table 3 and were first analyzed by nonparametric tests of difference of response rate under hypothetical and real conditions in order to assess the presence of hypothetical bias. The responses are split by the two conditions: the type of answer (Yes or No) and the nature of the payments condition (hypothetical versus real). A χ^2 test for independence is employed separately for the private and for the public provision cases and results are reported in table 4. In the case of public provision χ^2=1.32 and with one degree of freedom we cannot reject the null hypothesis of independence of the response rate from the hypothetical or real nature of the choice at any level lower than 25%. For the private provision case χ^2 = 3.84, hence the null hypothesis is rejected at 5% confidence level.[1]
If we take the response rate of the real setting as a proxy of unbiased behaviour then we observe that in the private provision case there is a prevalence of No answers over Yes ones when the payment is real but not when it is hypothetical. Conversely, when information is provided publicly, a larger rate of Yes responses is observed in both settings. This evidence seems to support the common assumption that, for hypothetical provision of private goods, an upward bias of Yes responses should be

[1] Similar results are obtained if data from graduate students are eliminated (tables 5 and 6). However, in this case the hypothesis of independence of the response rate from the hypothetical or real setting is rejected only at the 7% significance level in the case of private provision of information.

observed (Carson & Grove 2007). In hypothetical surveys on the provision of public goods potential buyers would untruly answer “Yes” so as to signal the provider a larger collective benefit which would in turn enhance the likelihood of provision. However, we note that in this experiment the good was quite a peculiar one. Information was to be used in a one-shot game and so no such effect can arise as the game was not repeated. Therefore, no clear indication on the existence or direction of the bias was expected in this case. The slight majority of Yes responses observed in the hypothetical private setting might be simply due to the absence of budget constrains or other effects due to the desire of safe-guarding personal reputation in the group or to warm glow effects.

A further analysis was carried out by splitting the data across locations. Significant differences in response rates arose only for the private real settings where subjects in Florence show a higher rate of Yes than in the other two locations (tables 7 and 8). Overall the public treatment seems to give rise to more uniform responses. This may relate to the fact that in the public provision the overall decision depends on preference of other people and possibly common expectation are formed with respect to those preferences.

Some socioeconomic and risk attitude characteristics of the sample are illustrated in table 9. Although there does not appear to be major differences in these characteristics across treatments two probit models have been estimated to investigate the possible impact of demographics variables on the response rate in both the private and public provision treatments. Following Cummings et al. (1997), a dummy variables is added to the regressors to capture the hypothetical vs real setting of the experiment (HYPOTH). The results of the probit analysis are reported in table 10.

The overall regression models are significant at the 5-6% level with only the undergraduate dummy showing individual statistical significance. However, in the private provision the hypothetical dummy is significant at the 10% level. These results are not dissimilar from those obtained by Cummings et al. (1997) for a public good but in that case only the treatment dummy (real vs hypothetical) was significant and the authors concluded that the null hypothesis of no impact of hypothetical setting on the response rate should have been rejected. Conversely, in our case the undergraduate dummy in both equations, and to a lesser extent the risk loving measure in the public provision, show considerable statistical significance. Moreover, in the private setting there appear to be some differences, although only at the 10% level of significance, between the hypothetical and real treatments.

To take into account possible inequality in covariate effects and heteroskedasticity across experimental treatments, a sequential test on the joint hypothesis of equality of parameters and error variance is performed following the rescaling approach proposed by Haab et al. (1999)[2]. The analysis is performed separately for both the private and the public provision of information.

As we are particularly interested in the difference between hypothetical and real treatments, the null is equality of both regression and scale parameters in the two conditions: H3 :

$$H_3 : \begin{pmatrix} \beta_{re} = \beta_{hy} \\ \sigma_{re} = \sigma_{hy} \end{pmatrix}$$

[2] For a critical assessment of the Haab procedure see Harrison (2006).

Since the scale parameters are unidentifiable under the two different conditions a sequential testing procedure is adopted following Haab et al. (1999). Firstly, the equality of regression parameter is tested: $H_4 : (\beta_{re} = \beta_{hy})$

then, if the previous test fails to be rejected we proceed with testing the equality of variances:

$H_5 : (\sigma_{re} = \sigma_{hy})$

H4 implies that the scale parameters σ_{re} and σ_{hy} can take on any values. To deal with inequality in the scale parameters σ_{hy} is normalized to 1 and a relative scale factor μ is defined as the ratio of the scale parameter for the private treatment on the scale parameter for the hypothethical treatment data ($\mu = \sigma_{re}/\sigma_{hy}$). The relative scale factor is estimated through a one dimensional grid search over possible values of μ in order to maximise the likelihood of the probit on the data matrix:

$$X = \begin{pmatrix} X_{re} \\ \mu X_{hy} \end{pmatrix}$$

Once a value for μ is obtained, estimation under H4 is achieved by estimating a probit on the data matrix scaled with the likelihood maximizing μ. The alternative to H4 being inequality of parameters with no restriction on scale, it is estimated running independent probit on the public and private treatment data. A standard likelihood ratio test is then employed to test for H4. Tables 11 and 12 provide the necessary data respectively for the private and public provision cases. Columns 2 to 5 are associated with the rescaling procedure. Columns 2 presents the estimates of the pooled sample without the treatment variable HYPOTH. Columns 3 and 4 are the estimates of the independent probit over the subsamples of the Real and Hypothetical treatment data. Column 5 shows the estimates of the probit estimated over the scaled sample under H4 with the optimal μ equal to 0.32 for the private and 0.50 for the public provision of information. The unrestricted log-likelihood for the LR test is given by the sum of the log likelihoods of the independent probits:

$LR1 = -2 \times (L_{scaledsample} - (L_{hy} + L_{re}))$

$LR1_{private} = -2 \times (-180.25 - (-88.95 - 87.56)) = 8.20$ (p=0.15)

$LR1_{public} = -2 \times (-172.46 - (-86.01 - 85.77)) = 2.61$ (p=0.76)

The test yields a value lower than the value of χ^2 statistics at the 5% ($\chi^2_{.05,5} = 11.1$) level of significance. Thus H4 cannot be rejected at this level of significance. This means that the marginal effects of explanatory variables do not vary across condition of provision of the information when variance differences between the real and hypothetical treatments are taken into account. From our data there is no clear evidence that the hypothetical setting impacts on subjects' responses in both the private and public provision case. However, this result is weaker in the private provision case.

Having failed to reject H4, H5 (equality of variance hypothesis) can be tested comparing the models in column 2 and 5 of table 8. The log-likelihood function under the null is given by the pooled equation (column 2) while the unrestricted loglikelihood is obtained by estimating the scaled model (column 5). The likelihood ratio statistic can be written as:

$LR2 = -2 \times (L_{unscaledpooled} - L_{scaledsample})$,

$LR2_{private} = -2 \times (-180.61 + 180.25) = 0.73$ (p=0.39),

$LR2_{public} = -2 \times (-173.10 + 172.46) = 1.28$ (p=0.26),

which is to be compared with a critical value of ($\chi^2_{(.05,1)} = 3.84$). Even in this case we fail to reject the null hypothesis of equality of variances between the hypothetical and the real treatment subsamples. Therefore, H3 can not be entirely rejected. The

results of our analysis suggest that the distribution of Yes responses under the hypothetical setting of information is flatter than that for private provision, although not significantly so.
However, when the original model is re-estimated with scaled data (table13) the effect of the hypothetical dummy variable is significant in the private provision case while it is not significantly different from zero in the public provision sub-sample. This result is not unexpected as a closer examination of tables 11 and 12 reveals that the intercept of the real provision equation (column 3) is not significantly different from zero (z = 0.21) while the intercept of the hypothetical sample equation (column 4) shows a value of the z-statistic close to 2.[3] Looking back to the rescaled equation, two more variables appear to be significant for the private provision case: the undergraduate dummy (UNDERGR) and income (HOU_IN_COR). The former shows a negative impact on the probability of obtaining "Yes" responses in the valuation question while the income coefficient has the expected positive sign. Conversely, in the public provision case no statistically significant coefficient is obtained except for the undergraduate dummy that is significant at the 7% level.

5 Results from verbal protocol analysis

In order to shed some light into the motivations that underlie the pattern of responses described in the previous section a content analysis was performed on the transcript of the thought list boxes described in section 3. Table 14 illustrates the percentage of subjects participating in the different treatments that mentioned certain themes related to the task of deciding whether to vote Yes or No in the public provision treatments or whether to accept to buy the information in a sealed envelope in the private treatment.
The main codes emerged during the content analysis (all mentioned by more than 10% of the participants in at least one treatment) are described below:
1. *What others do*: these are thoughts respondents had about the behaviour of other participants. For example:
"Is the weather forecast correct? What did the others choose? Why did the others want the forecast?" [Hamilton, Real, Public provision]
These thoughts hint to a social dimension of the choice task in contrast with the assumed single agent maximisation problem. There is a slight predominance of these thoughts in the public provision setting.
2. *The rules of the game*: under this heading we collected all thoughts about the rules of the game (such as payments, voting rules, provision in sealed envelop of information). One issue raised by some participants was a sort of majority dictatorship problem with the ballot vote in the public provision:
"The fee should be paid only by those who were in favour of the forecast service provision" [Firenze, Real, Public provision]
Apparently, these issues are found more often in the hypothetical rather than in the real settings.
3. *Winnings as positional goods*: these are considerations among the opportunity to obtain larger gains than the other participants by acquiring the information message. In other words winnings are regarded as positional goods. For example:

[3] This suggests a possibility that the constants could be excluded from the effects of the scale factor as in Swait and Louviere (1993).

"Whilst I felt the information would be important for my predictions I felt I would lose any advantage if everyone had the information. Better that no-one has it than being forced to have to pay for it." [Hamilton, Hypothetical, Private provision]

4. *How others affect me*: the code is about how the behaviour of others affect the participant either by determining his/her gains or by giving rise to imitative behaviour (herding). The influence of others on participants gains may be illustrated by the following quotation:

If less than 50% people vote 'yes', then there will be no money to earn, because the chance is 50:50 and I don't have to pay for it. [Hamilton, Real, Public provision]

While an example of imitative behaviour is given by:

I think that most people will go for having it so I will too [York, Real, Public provision].

Not surprisingly, these considerations are more frequent in the public provision setting as they naturally relate to a collective decision context.

5. *Altruism, justice and cooperation*: thoughts about moral and ethical issues are gathered under this heading. For example:

"It is fair that everybody pays for the information message because the message could prove useful to everybody" [Firenze, Real, Public provision]

"I was thinking for myself it just so happened that it would now benefit everyone else." [York, Hypothetical, Public provision]

"It would be easier to make decisions if some information was available. Without the information there is 50% chance of success but with the information there is 75% chance of success. The information per person would be cheaper [when purchased publicly] than if purchased individually" [Hamilton, Hypothetical, Public provision].

This kind of considerations arises mostly in the ballot vote and in the hypothetical treatments.

6. *Risk*: this theme is about subjects referring to gambling or risk avoidance as reasons for their choices. For example:

The weather forecast has 75% accuracy. This is a pretty good accuracy, comparatively. So, I would like to use 2.75 pounds to 'gamble' in a rather reasonable way. [York, Real, Private provision]

Real treatments seem to convey this kind of consideration more frequently.

7. *Probability computation*: this theme is about engaging in some sort of more or less elaborate computation of the value of the information message (for example trying to sorting out what the message would be worth for a risk neutral subject):

"I thought that 2.75 was good value for a 75% chance of winning 10 pounds. Although now I look back and realise its not and its relative to a 50% chance of winning 10.00 pounds and paying nothing." [York, Hypothetical, Public provision]

8. *Budget*: this refers to considerations about participants' own income or reference to the incentive given to the subject to make him/her participate in the experiment. For example:

"I said: I am here and they give me 10 euro and I don't have to pay nothing. Never mind if I get 7.25 euro never but I get more chance to reach 17.25 euro" [Firenze, Real, Private provision]

9. *Role playing as farmer*: these are thoughts about what a farmer should do with a weather forecast and how the information would improve the livelihood of farmers:

" To generate growth for the crop I want favourable conditions, acquiring as much information as possible increases my chances of a successful crop" [Hamilton, Real, Public provision].

A number of issues were recalled both in the hypothetical and the real setting under public provision. Namely the real vs. hypothetical conditions does not seem to discriminate across codes 1, 5, 7 and 9. These are codes that refer to interaction with others (1), interdependent utilities (5), rational optimizing behaviour (7) and scenario involvement (9). By contrast, under private provision the codes that are observed in both hypothetical and real settings and are mentioned by at least 5% of participants are: 1, 7, and 8. Apart from code 7 that is in common with the public provision case only code 1 and 8 do not discriminate across the real vs hypothetical treatments.

Most other codes, that is rules of the game, winning as positional good, altruism justice and cooperation and role-playing as farmer are mentioned more often in the hypothetical treatments while considerations about chance and risk are raised more frequently by those that participated in the real private provision treatments.
Not surprisingly, the public provision with its referendum mechanism prompted a number of thoughts about others, the feelings of being uneasy with a decision taken by the majority (rules of the game), possibility of cooperation and concern about others' welfare. All these themes are not related to the computational effort of sorting out whether the information message was a good bargain or not (indeed for a "rational" risk neutral subject it was not, but only few participants effectively noticed it). Such computational issues were slightly more frequent among participants in the private provision treatments.
More interestingly, under public provision of the information message a larger number of thoughts are shared across participants in the hypothetical and real referenda. This is not the case for the private provision where in the real settings all issues mentioned by at least 10% of participants pertain to the structure of the economic decision problem: budget constraints, probability computation and risk aversion/loving.
Summarising, it could be argued that in the private provision treatments the hypothetical setting triggers a number of thoughts that distract participants' attention from factors that might be reasonably expected to influence willingness to pay. Conversely, participants in the ballot vote are influenced by considerations about other people in both real and hypothetical settings. These results cast doubts about the empirical relevance of the theoretical basis of incentive compatibility for the experimental design illustrated in this paper. Rather, the observed consistency between hypothetical and real ballot vote outcomes seems to be attributable to this common background of thoughts and consideration, background that lacks in the private provision case.
The question that remains open is whether this behaviour is due to the particular nature of the good provided, or it is a more general feature of the type of provision. Certainly, deciding about an information message that can be used in a one shot game is a particular task and requires computational efforts. Moreover, a close referendum on small groups is likely to strengthen the perception of interrelationships across participants.
Notwithstanding, the thought list content analysis still shows that a peculiarity exists for the public provision and that this peculiarity is strengthened by the hypothetical context. Further research is needed to better understand these features of the referendum mechanism that seem to be completely unrelated with the suggestion that advisory referendum can be incentive compatible if they are perceived as consequential.

6 Conclusions

This research has implemented well established experimental economics protocols adapted so as to unambiguously test for incentive compatibility of the referendum format for public and private good provision in hypothetical and real markets, without potential confounding due to the nature of the good, which in all cases consisted of information messages.
Our results are mixed. On the one hand in the public provision treatment no significant difference was found in the response rate between hypothetical and real setting. On the other hand in the private provision treatment results show some

effect. Non parametric tests show significant differences between hypothetical and real setting with a larger rate of Yes responses in the hypothetical one. The probit regression provides borderline results. At the 15% level of significance we were able to reject the hypothesis of equal coefficients in the real and hypothetical sub-samples, thus partially supporting the Carson and Groves (2007) hypothesis. However, we feel that our findings cannot be simply transposed into the real world of CVM practice. The question that remains open is whether the observed behaviour is due to the particular nature of the good provided, or it is a more general feature of the type of provision. Certainly, deciding about an information message that can be used in a one shot game is a particular task and requires computational efforts. Furthermore running a referendum decision rule in small groups is likely to strengthen the perception of interrelationships across participants. In other words, our public provision setting was what Harrison (2006) names a “social choice microcosm” in which each participant makes choices that influence directly others. To what extent this helps providing a stronger parallelism between the hypothetical and the real context is a question that deserves further thought and investigation.

7 References

Becker, G.M., De Groot, M.H., Marschak, J. (1964). Measuring Utility by Single-Response Sequential Method, *Behavioral Science*, 9:226-32.

Cacioppo, J. T., Von Hippel, W., Ernst, J. M. (1999) Mapping cognitive structures and processes through verbal content, the thought-listing technique. *Journal of Consulting and Clinical Psychology*, 65(6):928-940

Camerer C. (1995), Individual decision making, in : Kagel J.H and Roth A.E (edts), *The Handbook of Experimental Economics*, Princenton University Press, Princenton, NJ, 587-704.

Carson, R.T. (2000). Contingent valuation: a user's guide, *Environment, Science and Technology*, 34:1413-1418.

Carson, R.T. and Groves T. (2007). Incentive and informational properties of preference questions, Environmental and Resource Economics, 37, 181-210..

Cummings, R. G.; Elliott, S.; Harrison, G.W., and Murphy, J. (1997). Are Hypothetical Referenda Incentive Compatible?, *Journal of Political Economy*, 105(3):609-621.

Cummings, R.G., Harrison, G.W., and Rutström, E. E. (1995). Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?, *American Economic Review*, 85(1), March, 260-266.

Currim, I. S. and Sarin R. K. (1989)., Prospect versus utility, *Management Science*, 35(5), 22-41.

Davis, D.D. and Holt, C.A. (1993). *Experimental Economics*. Princenton University Press, Princenton, NJ

Diamond, P.A., Hausman J.A. (1994). Contingent Valuation: is some number better than no number?, *Journal of Economic Perspectives*, 8(4):45-64.

Haab, T.C.; Huang, J.C., Whitehead, J.C. (1999). Are hypothetical referenda incentive compatible? A comment, Journal of Political Economy, 107(1), 186-196.

Harrison, G. (1986), An experimental test for risk aversion, Economics Letters, 21, 7-11.

Harrison, G. (2006), Experimental Evidence on Alternative Environmental Valuation Methods. Environmental & Resource Economics , 34: 125–162.

Harrison, G.W. and Rutstrom, E. (2002), Experimental evidence on the existence of hypothetical bias in value elicitation methods, in Plott C.R. and Smith V.L. (eds), Handbook of Experimental Economic Results, North-Holland, Amsterdam.

Hey J. and Orne C. (1994), Investigating generalizations of expected utility theory using experiments, Econometrica, 62(6), 1291-1326.

Hirshleifer, J. and Riley, J.G. (1992). The Analytics of Uncertainty and Information. Cambridge University Press, Cambridge UK.

Holt, C.A. and Laury, S.K. (2002). Risk Aversion and Incentive Effects, The American Economic Review, 92(5), 1644-1655.

Jaeger, S. (2005). Invited commentary on van Kleef et al.: selection and evaluation criteria, Food Quality and Preference, 16, 209-212.

Lawrence, D. (1999). The economic value of information, Springer Verlag, New York.

Mc Clelland, G.H., Schulze, W:D. and Coursey, D.L. (1993), Insurance for low-probability hazards: a binomial response to unlikely events, Journal of Environmental Economics and management, 19, 244-263.

Schkade, David A. and Payne John W., 1994, How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation, Journal of Environmental Economics and Management, 26, 88-109.

Svedsater, H. (2003). Economic Valuation of the Environment: How Citizens Make Sense of Contingent Valuation Questions, *Land Economics*, 79(1):122-135.

Swait J. and Louviere J. (1993). The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models, *Journal of Marketing Research*, 30(3):305-314.

Wendt, D. (1969), Value of information for decisions, *Journal of Mathematical Psychology*, 6, 430-443.

Table 1 The ten paired lottery- choice decision

Option A				Option B			
High payoff		Low payoff		High payoff		Low payoff	
value	p	value	p	value	p	value	p
6	10%	4.8	90%	11.55	10%	0.3	90%
6	20%	4.8	80%	11.55	20%	0.3	80%
6	30%	4.8	70%	11.55	30%	0.3	70%
6	40%	4.8	60%	11.55	40%	0.3	60%
6	50%	4.8	50%	11.55	50%	0.3	50%
6	60%	4.8	40%	11.55	60%	0.3	40%
6	70%	4.8	30%	11.55	70%	0.3	30%
6	80%	4.8	20%	11.55	80%	0.3	20%
6	90%	4.8	10%	11.55	90%	0.3	10%
6	100%	4.8	0%	11.55	100%	0.3	0%

Table 2 Illustrative example of the sequence of messages

Message	True state	Correctness
Dry	Dry	Yes
Wet	Wet	Yes
Dry	Wet	No
Dry	Dry	Yes
Wet	Wet	Yes
Wet	Wet	Yes
Dry	Dry	Yes
Wet	Dry	No

Table 3 Unconditional experimental responses

Count

Provision		Real	Hypothet	Total
Private	N	70	54	124
	Y	64	80	144
	Total	*134*	*134*	*268*
Public	N	53	60	113
	Y	79	67	146
	Total	*132*	*127*	*259*

Response rate

Provision		Real	Hypothet	Total
Private	N	52	40	46
	Y	48	60	54
	Total	*100*	*100*	*100*
Public	N	40	47	44
	Y	60	53	56
	Total	*100*	*100*	*100*

Table 4 Chi squares tests

Chi Square		Chi square	df	As, Sig.
	Private	3.84	1.00	0.05
	Public	1.32	1.00	0.25

Table 5 Unconditional experimental responses (undergraduates only)

Count

Provision		Real	Hypothet	Total
Private	N	62	47	109
	Y	50	62	112
	Total	*112*	*109*	*221*
Public	N	39	48	87
	Y	68	58	126
	Total	*107*	*106*	*213*

Response rate

Provision		Real	Hypothet	Total
Private	N	55	43	49
	Y	45	57	51
	Total	*100*	*100*	*100*
Public	N	36	45	41
	Y	64	55	59
	Total	*100*	*100*	*100*

Table 6 Chi squares tests (undergraduates only)

Chi Square		Chi square	df	As, Sig.
	Private	3.31	1.00	0.07
	Public	1.72	1.00	0.19

Table 7 Comparison across locations (response rate)

			Florence	York	Hamilton	Total
Private	Real	N	33	55	71	52
		Y	67	45	29	48
		Total	100	100	100	100
	Hypoth.	N	49	43	29	40
		Y	51	57	71	60
		Total	100	100	100	100
Public	Real	N	40	43	36	40
		Y	60	57	64	60
		Total	100	100	100	100
	Hypoth.	N	57	49	33	47
		Y	43	61	67	53
		Total	100	100	100	100

Table 8 Comparison across locations (Chi square tests)

		Chi sq.	df	p
Private	Real	13.16	2	0.01
	Hypoth.	3.88	2	0.14
Public	Real	0.51	2	0.78
	Hypoth.	4.47	2	0.10

Table 9 Sample Characteristics

Subject Characteristics	Sample				
	Real-Priv.	Hyp-Priv.	Real-Publ.	Hyp.-Pub.	Total
MALE	0.53	0.52	0.47	0.50	0.50
UNDERGR	0.84	0.81	0.81	0.83	0.82
RISK LOV	4.40	4.13	4.10	4.18	4.20
RIS_REGO	0.84	0.77	0.87	0.80	0.82
HOU_IN_COR	8838	9370	9795	8265	9075

Note: MALE= is coded (1) male (0) female; UNDERGR is coded(1) undergraduate, (0) others; RISK LOV is the number of risky choices (ranging from 0 to 10)in the lottery game, for subject that switched more than once the value is replaced by the average of the regular ones ; RIS REGO is coded (0) if subject made more than one switch in the lottery game (1) otherwise; HOU_IN_COR is the household per capita income (either in pounds or Euro depending on the country): missing valued were replaced by sample mean conditional to country where experiments took place and responsibility for tuition (self, parents, grants).

Table 10 Regression result using the probit model : Provision of information

	Private provision			Public provision		
Variable	Coefficient	z-Statistic	Prob.	Coefficient	z-Statistic	Prob.
CONSTANT	0.39	1.19	0.23	0.31	0.95	0.34
HYPOTH	0.28	1.77	0.08	-0.21	-1.32	0.19
MALE	0.05	0.33	0.74	0.21	1.31	0.19
UNDERGR	-0.49	-2.32	0.02	0.39	1.87	0.06
RISK_LOV	-0.05	-0.88	0.38	-0.09	-1.71	0.09
HOU_IN_COR	0.02	1.53	0.12	-0.01	-1.01	0.31
Mean dependent var	0.54			0.56		
S.D. dependent var	0.5			0.5		
Sum squared resid	63.74			61.2		
S.E. of regression	0.49			0.49		
Log likelihood	-179.03			-172.22		
Restr. log likelihood	-185.02			-177.42		
LR statistic (5 df)	11.97			10.39		
Probability(LR stat)	0.04			0.07		
Obs with Dep=0	124			113		
Obs with Dep=1	144			146		
Total obs	268			259		

Table 11 Rescaling procedure results: private provision of information

	Original estimates	Pooled sample	Real sample	Hypoth sample	Scaled Sample Scale fact.=.32
	(1)	(2)	(3)	(4)	(5)
C	0.389	0.557	0.096	0.840	0.416
	(1.190)	(1.786)	(0.210)	(1.904)	(0.959)
HYPOTH	0.277				
	(1.775)				
MALE	0.053	0.051	0.296	-0.125	0.203
	(0.331)	(0.319)	(1.307)	-(0.533)	(0.947)
UNDERGR	-0.495	-0.506	-0.530	-0.429	-0.641
	-(2.321)	-(2.382)	-(1.714)	-(1.447)	-(2.174)
RISK_LOV	-0.045	-0.052	-0.045	-0.044	-0.060
	-(0.876)	-(1.010)	-(0.650)	-(0.559)	-(0.900)
HOU_IN_COR	0.016	0.017	0.038	0.001	0.034
	(1.534)	(1.605)	(2.097)	(0.062)	(2.052)
Observation	268	268	134	134	268
ln(L)	-179.03	-180.61	-87.56	-88.95	-180.25

Note: t statistics are reported in parenthesis

Table 12 Rescaling procedure results: public provision of information

	Original estimates	Pooled sample	Real sample	Hypoth sample	Scaled Sample Scale fact.=.50
	(1)	(2)	(3)	(4)	(5)
C	0.315	0.215	0.150	0.292	0.291
	(0.954)	(0.671)	(0.308)	(0.661)	(0.691)
HYPOTH	-0.210				
	-(1.318)				
MALE	0.212	0.207	0.338	0.071	0.301
	(1.309)	(1.282)	(1.470)	(0.305)	(1.467)
UNDERGR	0.393	0.380	0.450	0.291	0.466
	(1.870)	(1.810)	(1.572)	(0.889)	(1.801)
RISK_LOV	-0.091	-0.092	-0.078	-0.099	-0.116
	-(1.713)	-(1.735)	-(0.885)	-(1.480)	-(1.593)
HOU_IN_COR	-0.011	-0.009	-0.010	-0.010	-0.011
	-(1.011)	-(0.886)	-(0.775)	-(0.533)	-(0.885)
Observation	259	259	132	127	259
ln(L)	-172.22	-173.10	-85.77	-86.01	-172.46

Note: z statistics are reported in parenthesis

Table 13 Regression result using the probit model : rescaled original estimates

	Private provision			Public provision		
Variable	Coefficient	z-Statistic	Prob.	Coefficient	z-Statistic	Prob.
CONSTANT	0.31	0.71	0.48	0.32	0.75	0.45
HYPOTH	0.78	2.16	0.03	-0.15	-0.58	0.56
MALE	0.21	0.96	0.34	0.30	1.48	0.14
UNDERGR	-0.62	-2.12	0.03	0.47	1.82	0.07
RISK_LOV	-0.05	-0.79	0.43	-0.12	-1.59	0.11
HOU_IN_COR	0.03	1.98	0.05	-0.01	-0.92	0.36
Mean dependent var	0.54			0.56		
S.D. dependent var	0.50			0.50		
Sum squared resid	63.74			61.23		
S.E. of regression	0.49			0.49		
Log likelihood	-177.90			-172.29		
Obs with Dep=0	124			113		
Obs with Dep=1	144			146		
Total obs	268			259		

Table 14 Thought list content analysis. Percentage of participants who mentioned a specific theme across experimental cells

	Private	Public

	Real	Hypoth	Real	Hypoth
What others do	8	6	12	9
The rules of the game	1	7	6	12
Winings as positional goods	1	6	2	10
How others affect me	3	3	13	6
Altruism, Justice and Cooperation	1	13	14	24
Risk	22	9	14	7
Probability calculus	38	34	27	24
Budget	22	15	15	7
Roleplaying as farmer	6	12	11	17
Total number of participants	100	100	100	100

Table 15 Thought list content analysis. Percentage of participants who mentioned a specific theme: Real vs. Hypothetical and Public vs. Private provision comparisons

	Real All	Hypoth All	Private All	Public All	All sessions
What others do	10	7	7	10	9
The rules of the game	3	9	4	9	6
Winning as positional goods	1	8	3	6	5
How others affect me	8	5	3	10	6
Altruism, Justice and Cooperation	8	18	7	19	13
Risk	18	8	15	11	13
Probability calculus	32	29	36	25	31
Budget	19	11	19	11	15
Roleplaying as farmer	8	15	9	14	11
Total number of participants	100	100	100	100	100