ÖREBRO UNIVERSITY
SWEDISH BUSINESS SCHOOL

Working Paper

2/2010

Scale sensitivity and question order in the contingent valuation method

Henrik Andersson
henrik.andersson@tse-fr.eu
Toulouse School of Economics
(LERNA, UT1, CNRS)
Toulouse
France

Mikael Svensson
mikael.svensson@oru.se
Department of Economics
Swedish Business School
Örebro University,
Sweden

ISSN 1403-0586

http://www.oru.se/esi/wps
Örebro University
Swedish Business School
701 82 Örebro
SWEDEN

Scale sensitivity and question order in the contingent valuation method

Henrik Andersson[1, *], Mikael Svensson[2]

[1] Toulouse School of Economics (LERNA, UT1, CNRS), Toulouse, France
[2] Dept. of Economics, Swedish Business School, Örebro University, Sweden

[*] To whom correspondence should be addressed. Corresponding address:
Toulouse School of Economics (LERNA), 21 all.de Brienne, Aile J.J. Laffont, 31042 Toulouse, France, e-mail: henrik.andersson@tse-fr.eu

Acknowledgements Financial support from Vägverket, VTI and the Swedish Civil Contingencies Agency is gratefully acknowledged. The usual disclaimers apply.

Abstract: This study examines the effect on respondents' willingness to pay to reduce mortality risk by the order of the question in a stated preference study. Using answers from an experiment conducted on a Swedish sample where respondents' cognitive ability was measured and where they participate in a contingent valuation survey it is found that scale sensitivity is the strongest when respondents are asked about a smaller risk reduction first ("Bottom-up" approach). This contradicts some previous evidence in the literature. It is also found that the respondents' cognitive ability is correlated with their answers being line with theoretical predictions. The latter being important for the validity of the answers. Hence, the results of this paper suggest that scale sensitivity is related to the order of the questions and to respondents' cognitive ability.

Keywords: Cognitive ability; Contingent valuation; Mortality risk; Order effect; Scale sensitivity
JEL-Codes: D80; I10; Q51

1. Introduction

Economists usually prefer revealed-preference (RP) to stated-preference (SP) methods when non-marketed amenities are to be evaluated. Preferences in RP studies are not only based on actual choices made by individuals, but it is also assumed that actual choices, compared to hypothetical choices made by respondents in SP studies, are made on a more informed basis. However, SP methods have an important role to play when knowledge among analysts about the alternatives and consequences decision makers face is limited, or when market data does not exists for the amenity of interest. One SP method that has been used to evaluate a wide range of non-marketed amenities is the contingent valuation method (CVM), a method where respondents are asked directly about their willingness to pay (WTP) for the amenity.

CVM has been under heavy criticism for being inadequate to measure individual preferences (Diamond and Hausman 1994). In addition to the criticism of hypothetical bias (List and Gallet 2001; Murphy et al. 2005), much of the criticism has been based on the lack of scale sensitivity found in many CVM studies.[1] This has particularly been found in studies on "non-use values", where respondents have been found to be willing to pay the same amount to save different number of wild species or to protect different amounts of other environmental resources (Kahneman and Knetsch 1992; Desvouges et al. 1993). Whereas the advocates of CVM argue that the lack of scale sensitivity is a result of bad survey design (Smith 1992; Carson and Mitchell 1995; Carson et al. 2001), its critics argue that the CVM is not capable of eliciting individual preferences (Desvouges, Johnson et al. 1993; Diamond and Hausman 1994; Kahneman et al. 1999). Recent research suggests, however, that a general dismissal of the CVM on the basis of scale insensitivity is unwarranted, and that scale insensitivity can be in line with economic theory, and to be expected also in well conducted studies (Heberlein et al. 2005). For instance, if respondents prefer to save 300 rather than 800 wolves (Heberlein, Wilson et al. 2005, p.16), a

[1] *Scale* and *scope* are used interchangeably in the literature to define the size of the good. We choose to use scale, since it dominates in the literature on mortality risk reductions.

standard scale test in a CVM-study would falsely reject that the WTP answers are valid estimates of the respondents' preferences.

When individuals are asked about their WTP for a reduction of their own mortality risk we expect their WTP to be increasing with the size of the risk reduction. Since the respondent are expected to prefer a larger risk reduction to a smaller one, scale sensitivity is a necessary condition for WTP estimates for risk reductions to be valid measures of respondents' preferences. Moreover, a necessary condition for valid estimates is also that WTP should be near proportional to the size of the risk reduction (Hammitt 2000). Empirical findings show, though, that respondents' WTP to reduce mortality risk is not scale sensitive in line with the validity requirements (Hammitt and Graham 1999). Hammitt and Graham (1999) in their review of 25 CVM-studies on mortality risk reductions found that WTP was scale sensitive in most studies, but in no study was it proportional to the risk reduction. Carson et al. (2001) suggest that eliciting individuals' preferences for reductions in the probability of death is especially cognitively constraining and recent research have found evidence which suggests that respondents who understand the scenario are more capable of giving answers in line with the theoretical predictions (Corso et al. 2001; Alberini et al. 2004; Andersson and Svensson 2008).

The presence of order effects have also been used to criticize CVM (Bateman and Langford 1996; Powe and Bateman 2003; Clark and Friesen 2008). Order effects imply that the WTP the respondents’ state for a good varies according to the order in which it appears in a survey relative to other goods. As an example, considering two different safety goods differing only with respect to the size of the mortality risk reduction (a “small” or a “large” reduction). If a respondent is asked to state his/her WTP for both risk reductions one after the other, the theoretical expectation is that WTP for the small reduction should not be different depending on if the respondent is asked this question before or after the large risk reduction question, given that the goods in each case are exclusive. If the smaller good is valued before the larger good, this may be referred to as a ”Bottom-up” approach, whereas if the larger good is valued before the smaller good this may be referred to as a “Top-down” approach (Carson and Mitchell 1995). Hence, if the WTP for the small risk reduction is significantly different in a “Bottom-up” compared to a “Top-down” approach, there are order effects in the study. The presence of order effects could be taken as evidence for the argument that respondents do not have well-defined preferences, instead these are created when respondents are confronted with the survey scenario (Slovic 1995).

In this paper we merge the literature on scale bias and order effects and the aims of the study are to examine: (1) if the order of questions affect the WTP for mortality risk reductions in SP studies, (2) if scale sensitivity is affected by using a “Bottom-up” or a “Top-down” approach, and (3) the effect of cognitive ability on potential order effects and scale sensitivity. Answers from a Swedish experiment are used in which respondents completed a simple test on cognitive ability before they conducted a CVM survey. The experiment was conducted in the fall of 2005 and in the CVM survey the respondents were asked about their WTP to reduce the mortality risk they face when riding the local bus. The respondents were asked two questions; half the sample stated their WTP for a smaller risk reduction before a larger one (“Bottom-up” approach), while the other half stated their WTP for the larger risk reduction first (“Top-down” approach).

The paper is structured as follows. In the following section we briefly present the theoretical model and describe empirical tests of scale sensitivity. Sections 3 and 4 contain a brief description of order effects and results in the literature, and a description of the survey and the data that was collected. The results are presented in section 5, which reveal that we cannot reject the null hypothesis of no order effects on the level of the WTP. However, we find evidence that suggests that the order of the question influence the scale sensitivity among respondents. Hence, we find no evidence that the level of WTP should be significantly affected by the order of questions, but that scale sensitivity may be. The latter being important for the validity of the respondents’ stated WTP. In section 6 the paper ends with a discussion and some concluding remarks.

2. Marginal willingness to pay to reduce mortality risk

Marginal WTP to reduce mortality risk is usually denoted the value of a statistical life (VSL), and is a measure of the population mean marginal rate of substitution between mortality risk and wealth (Dreze 1962; Schelling 1968). The decision individuals face can be illustrated with a state-dependent expected utility model (Rosen 1988). Let *w*, *p*, and $u_s(w)$, $s \in \{a, d\}$, denote wealth, baseline probability of death, and the state-dependent utilities, respectively, with subscripts *a* and *d* denoting survival and death. The survival lottery will then be given by :

$$EU_0(w, p) = pu_d(w) + (1-p)u_a(w). \tag{1}$$

For a marginal change of *p* we have the standard result that:

$$VSL = \left.\frac{dw}{dp}\right|_{\text{EU constant}} = \frac{u_a(w) - u_d(w)}{pu_d'(w) + (1-p)u_a'(w)}, \tag{2}$$

where prime denotes first derivative (Hammitt 2000). Under the reasonable assumptions (which are standard in the literature) that $u_a(w) > u_d(w),\ u_a'(w) > u_d'(w) \geq 0$ and $u_s''(w) \leq 0$ for $s \in \{a, d\}$, VSL is positive and increasing with *w* and *p* (Jones-Lee 1974; Weinstein et al. 1980).[2]

Equation (2) denotes "true" marginal WTP. In CVM-studies respondents are asked to state their WTP for a small but finite risk reduction, Δp. Let now w_0 and p_0 denote the initial wealth level and baseline risk in equation (1). WTP in equation (3) then defines the maximum amount an individual is willing to pay, remaining at the same utility level, to reduce the mortality risk by Δp, i.e.

$$EU_1 = (p_0 - \Delta p)u_d(w_0 - WTP) + (1 - p_0 + \Delta p)u_a(w_0 - WTP) = EU_0. \tag{3}$$

[2] The effect from the baseline risk (*p*) on VSL is often referred to as the "dead-anyway effect" (Pratt and Zeckhauser 1996).

As mentioned, a necessary (but not sufficient) condition for the WTP answers from CVM studies to be valid estimates of individuals' preferences for mortality risk reductions are that they are near-proportional to the size of Δp (Hammitt 2000).

Empirically scale sensitivity may be divided into "weak" and "strong" scale sensitivity (Corso, Hammitt et al. 2001). Weak scale sensitivity is fulfilled if WTP increases with the size of the risk reduction, while strong scale sensitivity refers to the situation where WTP increases near-proportionally to the magnitude of the risk reduction. In empirical applications scale sensitivity can be tested using either an internal ("within-sample") test, which consists of asking a respondent two or more WTP questions with different sizes of the risk reduction, or an external ("between-sample") test. The latter refers to varying the size of the risk reduction to different groups of respondents (where each respondent only answers one WTP question). Since we are interested in the order of questions and scale sensitivity in this paper the internal test is of most interest, and we require internal tests to answer our research questions.

3. Order effects

Order effects implies that the answer to the WTP question is affected by which order the questions are asked. The expected presence of order effects depends on if valuations are made from an inclusive or exclusive list (Bateman and Langford 1996). In inclusive lists, each subsequent good is described to be added to the previously valued good(s), while in exclusive lists goods are presented as alternatives to any other goods given in the list. Hence, in exclusive lists reference income, prices, quantity of goods consumed and utility level does not change across the valuation questions. Considering this, Bateman and Langford (1996) distinguish between theoretically expected "sequence effects" from inclusive lists due to income and substitution effects, and "order effects" from exclusive lists that are not consistent with theoretical predictions. Further, the literature has made a distinction between studies using advance disclosure or stepwise designs (Bateman et al. 2004). Advance disclosure refers to the situation where the respondent up front is being told about the different goods that are to be valued. In stepwise designs the respondent may first be asked to value good A, and only after that valuation being told now to value good B etc. In this paper we focus on and test for order effects in exclusive lists with advanced disclosure, where economic theory does not predict any significant order effects to be present.

Empirical evidence on order effects in exclusive lists includes two early studies by Boyle et al. (1990; 1993), who find some evidence for order effects, but not statistically significant in all cases. Powe and Bateman (2003) asked respondents for their WTP for a prevention program of salt-water flooding in a smaller or a larger scale. They found that the smaller amount of the program was valued significantly higher if asked first ("Bottom-up"), rather than if asked after the larger part ("Top-down"). There was also evidence that WTP for the larger good was higher in the "Bottom-up" approach. In the study by Bateman et al. (2004) they find significant order effects for a sample with stepwise designs, but no significant order effects in a sample with advance disclosure. For the smallest part of the good the mean WTP was highest in the "Bottom-up" approach, while the largest part of the good was highest in the "Top-down" approach.[3] Recently, Clark and Friesen (2008) test for order effects using an incentive-compatible Becker-DeGroot-Marschak mechanism for induced value goods, actual private goods and a private good being donated to charity. The find significant order effects for the private goods, such that both the smallest and largest good is valued higher in the "Bottom-up" approach. However, no order effect is found for the good being donated to charity. And they find no difference in scope sensitivity between the "Bottom-up" and the "Top-down" approach. Considering that Clark and Friesen (2008) report order effects also for the private good when elicited using incentive-

[3] Scope sensitivity is also tested by Bateman et al. (2004) and the results indicate scope sensitivity both in a within-sample (internal) test and in the between-sample (external) test.

compatible mechanisms may indicate that order effects is not something particularly associated with CVM studies, but rather a general phenomena also apparent in “real life decisions”.

4. The experiment – cognitive ability and a contingent valuation

The contingent valuation survey was part of an experiment conducted on 200 undergraduate students at Karlstad University in Sweden in the fall of 2005. The aim of the experiment was to examine how the respondents’ level of cognitive ability was correlated with WTP answers in line with the predicted scale sensitivity. A positive correlation was found between cognitive ability and scale sensitivity. This study focus on ordering effects on scale sensitivity and we refer to Andersson and Svensson (2008) for the analysis on cognitive ability and scale sensitivity.

Most of the students in the experiment were business majors (70 percent), but it also included students of economics, human resources, teaching and political science.[4] The students were recruited by being informed in class that they would be given the opportunity to take part in an experiment after the next lecture to provide valuable information for government authorities within the transport sector. They were informed that participation would be voluntary and that they would receive SEK 50 (ca. USD 7) as compensation for their participation.[5] One person supervised the whole experiment and we therefore do not expect heterogeneity in the responses due to undue influence exerted by different interviewers.

The individual attributes used as covariates in the regressions when analyzing order effects and scale sensitivity consist of the test score on the cognitive ability test and answers to follow-up questions about the respondents’ age, accident experience, etc. The test of cognitive ability consisted of 17 questions and was not in any way a complete test, but rather focused on skills in probabilities, syllogisms and computation. Hence, the test was a crude measure of cognitive ability but the questions used can be found in previous experiments and are similar to those used

[4] Since a preliminary analysis showed that the area of study had no significant influence on the WTP answers this analysis based on area of study is omitted.
[5] All prices in the paper are 2005 prices. USD 1 = SEK 7.48 (www.riksbanken.se, 01/27/06.)

in intelligence tests (Kahneman and Tversky 1972; Kahneman and Tversky 1983; Rabin 2002; Frederick 2005). Individuals often make decisions based on heuristics (short cuts), since mental short cuts lighten the cognitive burden of decision-making (Kahneman et al. 1982; Kahneman 2003), and the information from the test score is, therefore, valuable to test for a correlation between cognitive ability and order effects and scale sensitivity.[6]

In the CVM part of the experiment respondents were asked about their WTP to reduce the risk of being involved in a fatal bus accident. Bus fatalities constitute ca. one percent of all road accident fatalities in Sweden (SIKA 2005). This fatality risk was chosen for two reasons; familiarity and exogeneity. We assumed that the risk associated with using the bus would be both familiar and relevant to the students since many of the them need to use the bus to get to/from the city center of Karlstad and/or the campus at Karlstad University, which are located approximately 9 kilometers (ca. 5.6 miles) apart. Since the risk of riding a bus to large extend is related to circumstances out to the passenger’s control, e.g. the condition of the bus, the driver's behavior, and other elements of the traffic situation, we believe that we have less problem of a risk perceived as endogenous. That is, we believe that respondents are more likely to perceive the risk as exogenous compared with a scenario where they would have been asked about, e.g., the risk while driving a car, riding a bike, or smoking, since in these scenarios they can influence the risk by their own skills.

Before the WTP questions the respondents were informed about the overall mortality risk of a person in his/her 50s and of the average road-traffic fatality risk.[7] The risks were visualized using a grid containing of 10,000 whites squares where the appropriate number of squares had been blacked out to represent each risk. The visual aid was combined with verbal probability analog

[6] For a description of the test of cognitive ability and the questions used, see Andersson and Svensson (2008). The design of the CVM survey is available upon request from the authors.

[7] The baseline risk of this particular age group was used by both Persson et al. (2001) and Carlsson et al. (2004), and we decided to use it since we wanted a baseline risk other than the respondents’ own.

for the road-fatality risk, since this combination has proved to provide answers more consistent with standard economic theory (Corso, Hammitt et al. 2001). The probability analog said that the risk was equivalent to eight individuals on average dying annually in road-traffic accidents in a city of the size of Karlstad.

The first WTP question is illustrated in Figure 1. The open-ended format was chosen to avoid anchoring effects (Green et al. 1998) and in order to examine if scale bias is influenced by preference certainty, each WTP question was followed by a qualitative "certainty question", see, e.g. Blumenschein et al., (2008). The respondents were each faced with two WTP questions, one for the smaller risk reduction $\Delta p_S=4\times10^{-5}$, and one for the larger $\Delta p_L=6\times10^{-5}$. Half of the sample first stated their WTP the smaller risk reduction followed by a question on the larger risk reduction. The order was reversed for the other half. The second question was identical to the first one besides that Bus company B had been replaced by another company C with a different risk level. Since both samples were asked about the same risk reductions, the ratio between the risk reductions is equal for all respondents, i.e. 1.5.

[Figure 1 about here]

To avoid a framing effect by letting respondents compare new bus companies with an existing one, all were presented as new companies, but with one as the reference.[8] Moreover, to test for anchoring effects from the initial bus fare, half of the sample was told that the annual bus fare of the reference bus company was SEK 3,000, whereas the other half that it was SEK 4,200.[9] Hence, in all, the sample consisted of four subsamples based on question order and level of bus fare.

[8] The survey asks about the city council selecting from one of three bus companies that are distinguished by the levels of safety, when contracting out the service to a private firm. This corresponds to the actual procedure in the city of Karlstad as well as most other Swedish cities.
[9] The annual bus fare at the time of the survey in Karlstad was SEK 3,690

5. Results

The following sections contain the descriptive statistics and the tests for order effects and scale sensitivity. Tables 1 and 2 in the next section lists summary statistics and descriptions of variables used and show regression results on respondents' answers to their first WTP question, respectively. Section 5.2 then shows statistics and regression results related to the analysis of order effects and scale sensitivity.[10]

5.1 Descriptive statistics and willingness to pay regression

Many of the variables of Table 1 are self-explanatory but some may need further explanation. The variables *Weak sensitivity*, *Proportion* and *Absolute deviation* in Table 1 indicate how well, according to our hypothesis, respondents answered. *Weak sensitivity* is a dummy variable which takes the value one if respondents state a higher WTP for the larger risk reduction than for the smaller one. *Proportion* is a measure of the ratio between the WTP for the larger and smaller risk reductions, and is equal to 1.5 for fourteen percent of the sample and has a mean value of 1.58, which is close to proportionality. *Absolute deviation* is estimated as the absolute difference between the WTP ratio and risk-reduction ratio. For instance, since the ratio between the risk reductions is 1.5, a WTP ratio of 1.75 or 1.25 will mean an absolute deviation of 0.25.

[Table 1 about here]

The regression results in Table 2 are based on the respondents' stated WTP to the first question. Hence, the test of scale sensitivity is an external test since the regression only includes information about each respondent's WTP for one risk reduction. For the scale test we focus on the variable *Top-down*, which is a dummy equal to one for those respondents who answered the

[10] Since we use standard regression techniques we have not included a section where these techniques are described. For a description of the techniques, see any textbook on econometrics.

large valuation question first. The coefficient estimate for this variable is 0.42 and it is statistically significantly different from zero. A proportional WTP would result in a coefficient estimate equal to 0.405. A test shows that your estimate of 0.42 is not statistically significantly different from 0.405 (p-value = 0.91). Thus, we can reject the null hypothesis of no scale sensitivity but not the one of a proportional WTP. Regarding the other covariates, none of the coefficient estimates are statistically significantly different from zero.

[Table 2 about here]

5.2 Testing for order effects and scale sensitivity

In Table 3 WTP for the small and large risk reduction is displayed for the Bottom-up and Top-down samples. Further, the data is split with regards to the reference price in the survey (as either "low fare" or "high fare"). The prediction from economic theory is that WTP for each risk reduction should be equal in the "Bottom-up" and "Top-down" approach for the low fare, high fare and pooled sample, respectively. The results show that the WTP: (1) is higher with the "Top-down" approach, i.e. where the respondents value the larger risk reduction first, and (2) seems higher among those who were presented with the higher bus fare. However, none of these differences are statistically significant and we, therefore, cannot reject equality of WTP between the "Bottom-up" and "Top-down" approach or between the subsamples based on the level of the bus fare. Hence, based on these comparisons there are not statistically significant evidence in favour of any order or framing effects on the level of the WTP.

Regarding scale sensitivity we first examine external sensitivity, i.e. between groups. We find scale sensitivity for the first WTP answer, in line with the result in Table 2, and the pooled sample. For the second WTP answer grouped based on the bus fare we find no statistically significant difference in WTP for the small and large risk reduction, which would suggest an anchoring effect on the respondent's first WTP answer. The internal test of weak scale sensitivity reveals that the null hypothesis of no scale sensitivity can be rejected at the 1% level for all subsets. The ratios in column 4 are a measure of the WTP for the large risk reduction divided by the WTP for the small risk reduction (for each respondent). If there is no scale sensitivity or perfect proportionality the ratios should be equal to 1 and 1.5, respectively.[11] Our main interest in the ratios is to examine if there are any order effects. The results reveal that the ratios are statistically significantly different between "Top-down" and "Bottom-up" in the pooled sample and in the sample presented with the high bus fare. Hence, the empirical results suggest scale sensitivity is related to the order of the questions, with higher scale sensitivity in the "Bottom-up" approach.

[Table 3 about here]

In Table 4 we extend the analysis of order effects by including other covariates as summarized in Table 1. We run separate regressions on the log of the WTP for the small and the large risk reduction. The variable of interest is *Top-down*, which is expected to be zero based on predictions from economic theory. The test is external (between groups), and in line with the results in Table 3 we find no order effect. For both subsamples we find a positive coefficient for the "Top-down" approach which indicates a higher WTP if the respondent was asked about the larger risk reduction first. However, neither coefficient is statistically significantly different from zero. Moreover, it can also be seen that there is no effect (statistically significant) for the score on the cognitive test and WTP. We have also tested interactions between "Top-down" and the cognitive score to see if e.g. respondents with a low cognitive score who were asked about the larger risk

[11] One ratio, the Top-down approach with a low fare, is not statistically significantly different from one, which seems to contradict the statement in text about weak scale sensitivity. The reason for the different conclusions is that our tests of scale sensitivity are based on non-parametric tests.

reduction first are more prone to show order effects, but no such effects are found (and results not reported in Table 4).

[Table 4 about here]

In Table 5 a test of internal weak scale sensitivity is shown. The dependent variable (*Weak scale sensitivity*) is equal to one if the respondent valued the larger risk reduction higher and we run a binary probit model. In the regression on the full sample the coefficient for the variable “Top-down” is positive, but not statistically significant. Hence, we cannot reject the null hypothesis that weak scale sensitivity do not differ between the “Top-down” and the “Bottom-up” approach. Regarding other results, we find that cognitive ability is statistically significantly correlated with the probability of showing weak scale sensitivity. In column 3 and 4 the regression for weak scale sensitivity is run for the “Bottom-up” and “Top-down” approach subsamples, respectively. The results reveal that the effect from cognitive ability is higher in the “Top-down” approach. These results imply that weak scale sensitivity is related to respondents’ cognitive ability and can be interpreted as cognitive ability is more important when respondents are first asked about the larger risk reduction. Thus, the latter may imply that the cognitive task is harder when the first risk reduction is larger (“Top-down” approach). However, the difference is not statistically significant on conventional levels (p-value = 0.19). The only other significant covariates are *Risk help* and *Certain*, where the former is positive in the regression on the full sample and for the subsample ”Bottom-up” approach, and the latter negative in the subsample “Bottom-up” approach.

[Table 5 about here]

The results for strong scale sensitivity are shown in Table 6. In the full sample the deviation from proportionality is larger in the "Top-down" approach but the result is not statistically significant according to standard rules of thumb (p-value = 0.13). Regarding other results, again we find a statistically significant correlation between scale sensitivity and cognitive ability. The negative coefficient estimates suggest that the deviation from a proportional WTP is lower for respondents with a higher cognitive ability. We also find that the deviation from a proportional WTP is larger among women and lower among respondents with bus accident experience. Once again, looking at separate regressions for the two subsamples (column 3 and 4), the effect of cognitive is larger in the "Top-down" approach sample, even though the difference is not statistically significant on conventional levels (p-value = 0.16).

[Table 6 about here]

In Table 7 results from multinomial logit regressions on strong scale sensitivity are shown. Three regressions were run where the range of the interval of the base outcome is varied. For Strong scale sensitivity #1 where the base outcome is perfect proportionality the only variable that is statistically significant is *Female* among those whose WTP was more sensitive than predicted by theory. When the interval for the base outcome is widened in Strong scale sensitivity #2 and #3, two interesting effects can be observed: (1) those with a higher cognitive ability are less likely to state a WTP that is less sensitive than the relevant threshold (1.3 and 1.25) and (2) those who where first asked about the larger risk reduction ("Top-down") are more likely to state a WTP that is more sensitive than the relevant upper threshold (1.7 and 2). The latter result suggests that respondents' WTP are more likely to be too sensitive if they were asked about the larger risk reduction first, i.e. an order effect on the scale sensitivity. Moreover, respondents with a higher income level are more likely to have a too sensitive WTP, whereas those with bus accident experience are less likely.

[Table 7 about here]

6. Discussion

This article examines three main research questions: (1) testing for order effects in a CVM study on the WTP for mortality risk reductions, (2) examining if the scale sensitivity in WTP for the two different risk reductions is affected by the order of the two WTP questions, and (3) examining the effect of cognitive ability on potential order effects and scale sensitivity.

Regarding the first research questions the descriptive statistics showed that the “Top-down” approach, i.e. asking for the larger risk reduction first, produces higher WTP for both the larger and the smaller risk reduction. However, when comparing mean estimates (Table 3) and examining order effects and controlling for other covariates in regressions (Table 4) we find that the order effect is not statistically significant. Hence, we cannot reject the null hypothesis of no order effects on the level of the WTP. We used advanced disclosure in our survey, and the results of no order effects using advance disclosure is in line with the results in Bateman et al. (2004). In contrast to our results, Powe and Bateman (2003) found higher values in the “Bottom-up” approach, while Bateman et al. (2004) found higher values for the smallest good in the “Bottom-up” approach using stepwise design.

For the second research question of scale sensitivity our main interest is the internal sensitivity. We first focus on external tests of scale sensitivity, though, which can be considered the standard approach. These tests revealed a WTP that was sensitive to the magnitude of the risk reduction (Tables 2 and 3) and that proportionality is not rejected (Table 2) when the respondents’ first answer was used. However, when the respondents’ second answer was used we were not able to reject the hypothesis of no scale sensitivity. Hence, the external tests suggest that the first answers produce results in line with theoretical predictions, but that the second answers may have been anchored on the respondents’ initial answer.

Regarding internal scale sensitivity we found that scale sensitivity was higher among those who were presented with the “Bottom-up” approach (Table 3). This difference was statistically significant in the pooled sample and in the sample that were presented with the high bus fare (Table 3). Regarding our first analysis of strong scale sensitivity, when we examine deviation from proportionality (Table 5), we found a larger deviations from proportionality among respondents in the “Top-down” sample. Our second analysis of strong scale sensitivity was based on multinomial logit regressions where the band width of the deviation of proportionality was altered (Table 7). Among those who stated a WTP higher than proportional (i.e. too sensitive) we found when we increased the band width that respondents from the “Top-down” sample were more likely to be among those showing a too high scale sensitivity. Hence, there seems to be some evidence that the “Bottom-up” approach gives answers more in line with theoretical predictions. This suggests that scale sensitivity is indeed related to the order of the question.

Regarding cognitive ability and scale sensitivity we found that a higher score on the cognitive ability test was associated with a higher likelihood of showing scale sensitivity according to theoretical predictions (Andersson and Svensson 2008). Moreover, the results suggested that higher cognitive ability was more important in the “Top-down” compared with the “Bottom-up” approach, indicating that it is more cognitively demanding to give consistent answers in a “Top-down” approach. The differences between the samples were, though, not statistically significant on conventional levels; the test of difference between samples showed a p-value equal to 0.19 and 0.16 for weak scale and deviation from proportionality, respectively. In the multinomial logit, the effect of cognitive ability was statistically significant when the band width was increased and suggested that respondents with a higher cognitive ability were less likely to state a too insensitive WTP. Further, the effect increased with allowed band width.

Regarding some other results we found that respondents who stated to be certain of their answers did not state a lower WTP. This is interesting since some recent papers have argued that certainty is likely to reduce WTP and reduce the problem of positive hypothetical bias in SP studies (Blumenschein et al. 2001; Blumenschein, Blomquist et al. 2008). We find no evidence for this. More so, in some regressions we find that certainty is associated with a lower likelihood of showing weak scale sensitivity. Finally, worth mentioning is that bus mortality risk is not a pure private good, and there is a risk that values also represent paternalistic altruism (safety orientated) or that respondents answer strategically when they state their WTP for that kind of risk reduction. Note that, since we conducted internal tests on scale sensitivity, our findings should not be affected by potential strategic bias, altruism, or whether stated WTP also reflects preferences to reduce injury risk (the respondent's answers should be affected to the same relative degree).

To summarize, the results of this paper suggest that scale sensitivity is related to the order of the questions and to respondents' cognitive ability. There are some indications that results are more in line with theoretical predictions using the "Bottom-up" approach. Even if we find statistically significant evidence of order effects and scale sensitivity, other results are not statistically significant on conventional significance levels. It should be kept in mind, though, that even if the sample was larger than is usually the case for experiments, the analysis was carried out based on a sample smaller than is usually the case for CVM surveys. Therefore, this paper contributes to previous research within the area of preference elicitation and show that respondents' understanding of the scenario is important for valid answers and it adds to the literature on order effects by showing that respondents ability to understand the scenarios may explain part of these effects.

References

Alberini, A., Cropper, M., Krupnick, A. and Simon, N. (2004). "Does the value of a statistical life vary with age and health status? Evidence from the US and Canada," *Journal of Environmental Economics and Management* 48(769-792.

Andersson, H. and Svensson, M. (2008). "Cognitive Ability and Scale Bias in the Contingent Valuation Method," *Environmental and Resource Economics* 39(481-495.

Bateman, I., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G.L. (2004). "On visible choice sets and scope sensitivity," *Journal of Environmental Economics and Management* 47(71-93.

Bateman, I. and Langford, I.H. (1996). "Budget-constraint, temporal, and question-ordering effects in contingent valuation studies," *Environmental Planning A* 28(1215-1228.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N. and Freeman, P. (2008). "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment," *Economic Journal* 118(114-137.

Blumenschein, K., Johannesson, M., Yokoyama, K.K. and Freeman, P. (2001). "Hypothetical versus real willingness to pay in the health care sector: results from a field experiment," *Journal of Health Economics* 20(3), 441-457.

Boyle, K.J., Reiling, S.D. and Phillips, M.L. (1990). "Species substitution and question sequencing in contingent valuation surveys evaluating the hunting of several types of wildlife," *Leisure Sciences* 12(103-118.

Boyle, K.J., Welsh, M.P. and Bishop, R.D. (1993). "The role of question order and respondent experience in contingent valuation studies," *Journal of Environmental Economics and Management* 25(S80-S99.

Carlsson, F., Johansson-Stenman, O. and Martinsson, P. (2004). "Is Transport Safety More Valuable in the Air," *Journal of Risk and Uncertainty* 28(2), 147-163.

Carson, R.T., Flores, N.E. and Meade, N.F. (2001). "Contingent Valuation: Controversies and Evidence " *Environmental and Resource Economics* 19(2), 173-210.

Carson, R.T. and Mitchell, R.C. (1995). "Sequencing and Nesting in Contingent Valuation Surveys," *Journal of Environmental Economics and Management* 28(2), 155-173

Clark, J.E. and Friesen, L. (2008). "The causes of order effects in contingent valuation surveys: An experimental investigation " *Journal of Environmental Economics and Management* 56(2), 195-206.

Corso, P.S., Hammitt, J.K. and Graham, J.D. (2001). "Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation " *Journal of Risk and Uncertainty* 23(2), 165-184.

Desvouges, W.H., Johnson, F.R., Dunford, R.W., Hudson, S.P. and Wilson, K.N. (1993). "Measuring Natural Resource Damage with Contingent Valuation: A Test of Validity and Reliability", In J. A. Hausman (ed), *Contingent Valuation: A Critical Assessment.* Amsterdam: North-Holland.

Diamond, P.A. and Hausman, J.A. (1994). "Contingent Valuation: Is Some Number Better than No Number?," *Journal of Economic Perspectives* 8(4), 45-64.

Dreze, J.H. (1962). "L'utilité sociale d'une vie humaine," *Revue Francaise de Recherche Opérationnelle* 22(

Frederick, S. (2005). "Cognitive Reflections and Decision Making," *Journal of Economic Perspectives* 19(4), 25-42.

Green, D., Jacowitz, K., McFadden, D.L. and Kahneman, D. (1998). "Referendum contingent valuation anchoring and willingness to pay for public goods," *Resource and Energy Economics* 20(2), 85-116.

Hammitt, J.K. (2000). "Evaluating Contingent Valuation on Environmental Health Risks: The Proportionality Test," *AERE Newsletter* 20(1), 14-19.

Hammitt, J.K. (2000). "Valuing Mortality Risk: Theory and Practice," *Environmental Science & Technology* 34(8), 1396-1400.

Hammitt, J.K. and Graham, J.D. (1999). "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?," *Journal of Risk and Uncertainty* 18(1), 33-62.

Heberlein, T.A., Wilson, M.A., Bishop, R.C. and Schaeffer, N.C. (2005). "Rethinking the scope test as a criterion for validity in contingent valuation " *Journal of Environmental Economics and Management* 50(1), 1-22.

Jones-Lee, M.W. (1974). "The Value of Changes in the Probability of Death or Injury " *Journal of Political Economy*, 82(4), 835-849.

Kahneman, D. (2003). "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93(5), 1449-1475.

Kahneman, D. and Knetsch, J.L. (1992). "Valuing public goods: The purchase of moral satisfaction," *Journal of Environmental Economics and Management* 22(1), 57-70.

Kahneman, D., Ritov, I. and Schkade, D. (1999). "Economic preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues," *Journal of Risk and Uncertainty* 19(1-3), 203-235.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgement under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kahneman, D. and Tversky, A. (1972). "Subjective Probability: A Judgement of Representativeness," *Cognitive Psychology* 3(3), 430-454.

Kahneman, D. and Tversky, A. (1983). "Can irrationality be intelligently discussed?," *Behavioral and Brain Sciences* 6(509-510.

List, J.A. and Gallet, C.A. (2001). "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? ," *Environmental and Resource Economics* 20(3), 241-254.

Murphy, J., Allen, G.P., Stevens, T.H. and Weatherhead, D. (2005). "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation " *Environmental and Resource Economics* 30(3), 313-325.

Persson, U., Norinder, A., Hjalte, K. and Gralén, K. (2001). "The Value of a Statistical Life in Transport: Findings from a New Contingent Valuation Study in Sweden " *Journal of Risk and Uncertainty* 23(2), 121-134.

Powe, N.A. and Bateman, I. (2003). "Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs," *Ecological Economics* 45(255-270.

Powe, N.A. and Bateman, I.J. (2003). "Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs," *Ecological Economics* 45(255-270.

Pratt, J.W. and Zeckhauser, R.J. (1996). "Willingness to Pay and the Distribution of Risk and Wealth " *Journal of Political Economy* 104(4), 747-763.

Rabin, M. (2002). "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117(3), 775-816.

Rosen, S. (1988). "The Value of Changes in Life Expectancy," *Journal of Risk and Uncertainty* 1(3), 285-304.

Schelling, T.C. (1968). "The Life You Save May Be Your Own", In S. B. Chase (ed), *Problems in Public Expenditure Analysis.* Washington D.C.: The Brookings Institution.

SIKA (2005). Vägtrafikskador.

Slovic, P. (1995). "The construction of preferences," *American Psychologist* 50(364-371.

Smith, V.K. (1992). "Arbitrary values, good causes, and premature verdicts," *Journal of Environmental Economics and Management* 22(1), 71-89.

Weinstein, M.C., Shepard, D.S. and Pliskin, J.S. (1980). "The Economic Value of Changing Mortality Probabilities: A Decision-Theoretic Approach," *Quarterly Journal of Economics* 94(2), 373-396.

Figures

Figure 1 Willingness to pay question with small risk reduction and low annual bus fare

We would first like to know how much you are willing to pay to travel with Bus company B instead of Bus company A. The risks of accidents with fatal outcome for Bus company A and B are:

- *Bus company A:* Risk = 10 per 100,000
- *Bus company B:* Risk = 6 per 100,000

An annual pass with Bus company A costs SEK 3,000 (SEK 250 × 12).

What is the maximum amount you are willing to pay **more** per year to travel with **Bus company B** instead of **Bus company A**?

The maximum amount I am willing to pay more is . . . . SEK.

Are you definitely sure, probably sure or unsure regarding your answer (mark with an x)?

□ Definitely sure □ Probably sure □ Unsure

Tables

Table 1 Description of dependent and explanatory variables

Variable	**Description**	**Mean (Std. Dev.)**	**(Std. Dev.)**
Weak scale sensitivity	Dummy variable coded as one if respondent shows scale sensitivity by $WTP_{Large\ \Delta p}>WTP_{Small\ \Delta p}$	0.66	(0.48)
Proportion	Ratio for the respondent's WTP for the large and the small risk reduction	1.58	(0.93)
Absolute deviation	Deviation in absolute terms from linear proportionality	0.49	(0.79)
Top-down	Dummy for sub-sample that answered the larger risk reduction as the first WTP question	0.50	(0.50)
Low fare	Dummy for sub-sample where initial bus fare is lower	0.50	(0.50)
Female	Dummy equal to one if respondent is female and zero otherwise	0.54	(0.50)
Age	Age of the respondent	23.01	(3.94)
Income	Income based on medians from 8 different income categories (2005 prices, USD 1 = SEK 7.48)	6,836	(2,887)
Cognitive	Score on the cognitive test with a max of 17. (The lowest and highest score was 4 and 16.)	9.83	(2.12)
Bus	Dummy equals to one if respondents frequently travels by bus and zero otherwise	0.55	(0.50)
Bus accident	Dummy equal to one if respondent has personal experience of a bus traffic accident and zero otherwise	0.12	(0.32)
Risk help	Dummy variable coded as one if respondent stated that risk illustration was helpful and zero otherwise	0.24	(0.43)
Certain	Dummy equal to one if respondent stated definitely or probably sure in the WTP certainty question	0.76	(0.43)

Notes: Number of observations is 185.

Table 2 Regression analysis on 1st WTP answer

	Dependent variable: ln(WTP Q1)	
Variable	Coefficient	(Std. Err.)
Top-down	0.42***	(0.16)
Low fare	-0.21	(0.16)
Female	-0.08	(0.17)
ln(Age)	0.07	(0.16)
ln(Income)	-0.17	(0.15)
ln(Cognitive)	0.22	(0.33)
Bus	0.02	(0.17)
Bus Accident	0.07	(0.28)
Risk Help	-0.24	(0.19)
Certain	0.28	(0.18)
Intercept	6.28***	(1.25)
N	185	
R^2	0.08	

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Test of proportionality, i.e. coefficient for Top-down = 0.405, p-value = 0.91.

Table 3 Testing for order effects

	Mean WTP (std. dev.)			
	Small risk reduction	Large risk reduction	Ratio (std. err.)	N
Bottom-up $_{\text{High fare}}$	534 (686)	771 (959)	1.44*** (0.06)	48
Top-down $_{\text{High fare}}$	820 (1622)	1061 (1687)	1.29** (0.12)	46
Bottom-up $_{\text{Low fare}}$	562 (756)	707 (826)	1.26*** (0.07)	46
Top-down $_{\text{Low fare}}$	628 (923)	722 (828)	1.15 (0.16)	46
Bottom-up $_{\text{Pooled sample}}$	548 (719)	737 (889)	1.35*** (0.05)	94
Top-down $_{\text{Pooled sample}}$	724 (1316)	891 (1332)	1.23** (0.09)	92

Notes: (1) Framing effect of different bus fares not statistically significant. (2) External weak scale sensitivity statistically significant at 10% level or higher for pooled sample and for 1st WTP question for high and low fare subsets. (3) No statistically significant difference in WTP in external comparison (between) for same risk reduction. (4) Internal test of weak scale sensitivity statistically significant at 1% for all subsets. (5) Ratio statistically significantly different between Bottom-up and Top-down in high fare and pooled subsets at 10% level. (Tests of difference carried out with the Mann-Whitney test.)

Table 4 Regression analysis testing for order effects

	Dependent variable: Ln(WTP Large risk reduction)		Dependent variable: Ln(WTP Small risk reduction)	
Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
Top-down	0.08	(0.16)	0.05	(0.17)
Low fare	-0.25	(0.16)	-0.19	(0.18)
Female	-0.06	(0.17)	-0.21	(0.18)
Ln(Age)	0.02	(0.18)	0.09	(0.17)
Ln(Income)	-0.11	(0.16)	-0.20	(0.15)
Ln(Cognitive)	0.34	(0.34)	-0.17	(0.35)
Bus	0.06	(0.17)	-0.13	(0.18)
Bus Accident	0.06	(0.28)	0.10	(0.31)
Risk Help	-0.19	(0.19)	-0.21	(0.22)
Certain	0.19	(0.18)	0.23	(0.18)
Intercept	6.28***	(1.25)	7.79***	(1.59)
N	185		185	
R^2	0.04		0.04	

Notes: *** p<0.01, ** p<0.05, * p<0.10
Robust standard errors in parentheses.
Test of difference in coefficient estimates of Top-down, p-value = 0.85

Table 5 Test of weak scale sensitivity: Probit regression, dependent variable “Weak scale sensitivity”

Variables	Full sample	”Bottom-up” approach sample	”Top-down” approach sample
Top-down	0.10	-	-
	(0.19)		
Cognitive	0.15***	0.10*	0.23***
	(0.04)	(0.06)	(0.07)
Small	-0.10	-0.19	-0.06
	(0.20)	(0.30)	(0.30)
Female	0.13	0.03	0.30
	(0.20)	(0.28)	(0.29)
Age	-0.05	-0.04	-0.05
	(0.03)	(0.04)	(0.05)
Income	$5 \cdot 10^{-5}$	0.00	0.00
	$(4 \cdot 10^{-5})$	(0.00)	(0.00)
Bus	0.14	-0.12	0.39
	(0.20)	(0.29)	(0.29)
Bus accident	0.26	0.56	-0.20
	(0.33)	(0.54)	(0.45)
Risk help	0.43*	0.70*	0.31
	(0.24)	(0.39)	(0.34)
Certain	-0.24	-0.52*	-0.08
	(0.22)	(0.31)	(0.35)
Intercept	-0.45	0.34	-1.30
	(0.79)	(1.05)	(1.26)
N	198	99	99
R^2	0.08	0.10	0.13

Notes: *** p<0.01, ** p<0.05, * p<0.10.

Robust standard errors in parentheses.

Test of difference in coefficient estimates of Cognitive in Bottom-up and Top-down sample, p-value = 0.19

Table 6 Test of strong scale sensitivity, dependent variable: Ln(Absolute deviation)

Variables	Full sample	"Bottom-up" approach sample	"Top-down" approach sample
Top-down	0.05	-	-
	(0.03)		
Cognitive	-0.14**	-0.14**	-0.20**
	(0.05)	(0.07)	(0.09)
Small	0.02	-0.03	0.02
	(0.04)	(0.06)	(0.05)
Female	0.06**	0.11**	0.03
	(0.03)	(0.04)	(0.05)
Age	0.02	0.03	-0.12
	(0.02)	(0.02)	(0.12)
Income	0.04	0.03	0.05
	(0.03)	(0.03)	(0.06)
Bus	-0.01	0.01	0.06
	(0.04)	(0.05)	(0.05)
Bus accident	-0.05	-0.11**	0.01
	(0.04)	(0.05)	(0.05)
Risk help	0.01	-0.04	0.02
	(0.04)	(0.04)	(0.06)
Certain	-0.03	-0.10	0.05
	(0.04)	(0.06)	(0.05)
Intercept	0.47	0.65**	1.03
	(0.31)	(0.31)	(0.71)
N	184	93	91
R^2	0.06	0.12	0.09

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.
Robust standard errors in parentheses.
Test of difference in coefficient estimates of Cognitive in Bottom-up and Top-down sample, p-value = 0.16

Table 7 Test of strong scale sensitivity using multinomial logit

Variables	Strong scale sensitivity #1	Strong scale sensitivity #2	Strong scale sensitivity #3
		Low scale sensitivity	
Top-down	0.02	0.30	0.53
	(0.45)	(0.38)	(0.38)
Ln(Cognitive)	-1.23	-2.12^{**}	-2.86^{***}
	(0.97)	(0.84)	(0.87)
Small	0.27	0.36	0.73^{*}
	(0.43)	(0.38)	(0.39)
Female	0.43	0.17	0.39
	(0.45)	(0.38)	(0.39)
Ln(Age)	-0.36	1.95^{*}	1.34
	(0.89)	(1.18)	(1.10)
Ln(Income)	0.41	0.11	0.41
	(0.64)	(0.49)	(0.48)
Bus	-0.41	-0.50	-0.53
	(0.45)	(0.39)	(0.40)
Bus accident	0.17	-0.23	-0.18
	(0.71)	(0.54)	(0.53)
Risk help	-0.11	-0.23	-0.45
	(0.54)	(0.46)	(0.46)
Certain	0.56	0.22	0.16
	(0.49)	(0.41)	(0.42)
Intercept	0.94	-2.29	-1.84
	(5.03)	(4.74)	(4.66)
		High scale sensitivity	
Top-down	0.61	0.80^{**}	0.93^{**}
	(0.48)	(0.41)	(0.40)
Ln(Cognitive)	0.92	-0.54	-1.21
	(1.00)	(0.89)	(0.88)
Small	0.12	0.13	0.31
	(0.44)	(0.40)	(0.40)
Female	0.84^{*}	0.40	0.67^{*}
	(0.46)	(0.41)	(0.40)
Ln(Age)	-0.43	0.41	0.18
	(0.85)	(0.69)	(0.59)
Ln(Income)	0.85	1.12^{*}	1.26^{**}
	(0.69)	(0.57)	(0.57)
Bus	0.17	-0.17	-0.15
	(0.47)	(0.43)	(0.43)
Bus accident	0.05	-1.25^{*}	-1.17^{*}
	(0.70)	(0.72)	(0.71)
Risk help	0.26	0.05	0.02
	(0.54)	(0.47)	(0.47)
Certain	0.20	-0.27	-0.26
	(0.50)	(0.42)	(0.41)
Intercept	-8.45	-10.34^{*}	-9.77^{*}
	(5.83)	(5.66)	(5.56)
N	184	184	184
Pseudo-R^2	0.06	0.07	0.09

Notes: *** p<0.01, ** p<0.05, * p<0.10. Robust standard errors in parentheses. Multinomial #1: Base outcome Proportion=1.5, Multinomial #2: Base outcome Proportion$\in$ [1.3, 1.7], Multinomial #3: Base outcome Proportion$\in$ [1.25, 2).

Test of difference in coefficient estimates of Cognitive in #2 and #3 in Low scale, p-value = 0.03

Test of difference in coefficient estimates of Top-down in #2 and #3 in High scale, p-value = 0.34