Scope insensitivity in contingent valuation of complex environmental amenities

Knut Veisten[a], Hans Fredrik Hoen[b], Ståle Navrud[c] and Jon Strand[d]

[a] *Senior author (kve@toi.no)*

Department of Forest Sciences, Agricultural University of Norway, P.O. Box 5044, NO-1432 Ås, Norway
&
Institute of Transport Economics, P.O. Box 6110 Etterstad, NO-0602 Oslo, Norway

[b] *Department of Forest Sciences, Agricultural University of Norway, P.O. Box 5044, NO-1432 Ås, Norway*

[c] *Department of Economics and Social Sciences , Agricultural University of Norway, P.O. Box 5033, NO-1432 Ås, Norway*

[d] *Department of Economics, University of Oslo, P.O. Box 1095, NO-0317 Oslo, Norway*

May 2003

ABSTRACT

It has been argued that respondents in *contingent valuation* (CV) surveys, asked to value complex environmental amenities, will state willingness to pay (WTP) independently of the *scope* of the project. Such insensitivity to scope would be at odd with rational choice, and could therefore imply that CV is not a theoretically valid method for biodiversity valuation.

The scope test in the present CV study was over the composition of endangered species preservation. Respondents were split in four sub-samples facing different scopes of endangered species preservation. The design allowed for both external and internal scope tests. Furthermore, the tests were split according to elicitation format.

Of four external tests of insensitivity to scope, one was rejected, two gave mixed results, depending on either the type of test or elicitation format, and for the last one the null hypothesis could not be rejected. Of five internal tests, insensitivity to scope was rejected in three cases, one test gave mixed results, and one could not be rejected. Survey design features of the CV study, especially a fuzzy sub-group of endangered species, could explain the apparent insensitivity to scope observed.

Key words: contingent valuation, complex environmental amenities, passive use, scope insensitivity

1. Introduction

The contingent valuation (CV) method enables valuation of non-market goods by surveying a representative sample of the population and eliciting their *ex ante* willingness to pay (WTP) for specific changes (Bateman and Willis 1999, Bjornstad and Kahn 1996, Cummings *et al.* 1986, Mitchell and Carson 1989). This stated preference method has been assessed to produce estimates that have *construct validity* when applied to familiar public goods with predominant use values (Carson *et al.* 1996, Carson and Mitchell 1993, Cummings and Harrison 1995, Smith and Osborne 1996). There are more conflicting views regarding CV's validity when applied to goods with predominant passive-use (non-use) values (Bishop and Welsh 1992, Cummings and Harrison 1994, 1995, Hanemann 1994, Hausman 1993, Kopp 1992, Mitchell and Carson 1989, Smith 1992). As a tool for benefit cost analysis CV is more useful for passive-use values, since these goods in most cases only can be valued by stated preference.

Several studies have reported that respondents to CV surveys have stated the same WTP for goods that differed largely in scope or inclusiveness, when CV was applied to measure passive-use values (Boyle *et al.* 1994, Diamond *et al.* 1993, Kahneman 1986, Kahneman and Knetsch 1992, Kemp and Maxwell 1993, McFadden and Leonard 1993).[1] Insensitivity to scope conflicts with the rational choice behaviour assumed for agents in neo-classical economic theory. It follows from the assumption of non-satiated utility that WTP should increase with increasing numbers of a (normal) commodity, and WTP for a comprehensive good should be higher than for a subset of that good.

[1] A test of insensitivity to scope can be described as a *weak* test of economic theory, in the sense that falsification would eventually be the result of retaining, not rejecting, the null hypothesis (type I error). Kahneman and Knetsch (1992) term scope insensitivity *perfect embedding*, while e.g., Boyle *et al.* (1994) apply the term *part-whole bias* (see also Mitchell and Carson 1989).

The critics of CV claim that stated WTP for, e.g., preserving endangered species do not relate to economic choice behaviour but rather reflect *good cause dump* based on a *purchase of moral satisfaction* (Kahneman and Knetsch 1992). This is related to a theory of *mental accounting* (Tversky and Kahneman 1986, Thaler 1990), based on which it can be hypothesised that CV respondents will dump their mental account for environmental amenities on any project presented for them. Scope or size of the project would not matter.

Psychological explanations of scope insensitivity do not imply CV invalidation. Green and Tunstall (1999, p. 213) argue that observed scope insensitivity (part-whole bias, embedding) "is the result of asking questions which are essentially meaningless to the respondents because [of] false assumptions about the cognitions of the respondents". This position is close to that of, e.g., Carson and Mitchell (1993), arguing that apparent scope insensitivity is primarily due to flaws in survey design leading to *amenity misspecification* bias.

There are also explanations from economic theory. Rollins and Lyke (1998) argue that observed insensitivity to scope can result from diminishing marginal values. Successive quantities of, e.g., protected areas would receive ever positive but lower values per unit, such that the possibility of observing scope sensitivity would depend on the baseline scarcity of the resource. Income effects provide a related explanation. CV respondents have limited budgets or sub-budgets, whether these are mental or real, so their optimisation of spending on private and public goods is constrained (Randall and Hoehn 1993, 1996). Thus, even if the valuation is hypothetical, respondents are expected to limit totally stated WTP to their ability to pay and to account for an executed hypothetical purchase when asked to value another good.

Indeed, the scope sensitivity issue remains controversial. The NOAA *blue ribbon panel* on contingent valuation concluded that the observation of insensitivity to scope (*perfect embedding*) was "perhaps the most important internal argument against the reliability of the CV approach" (Arrow *et al.* 1993, p. 4607). Since then important literature assessments and a meta-analysis have concluded that CV respondents have shown sensitivity to scope. Smith and Osborne (1996) found scope-sensitivity in a meta-analysis pooling WTP estimates for visibility improvements at U.S. national parks from five CV studies. However, visibility represents a *non-consumptive use* value, and thus it can still be claimed that insensitivity to scope in CV of *passive use* remains problematic. E.g. Carson (1997) assesses several CV studies and finds that most of these show scope sensitivity, even when based on goods with an assumedly important passive-use component. But, Svedsäter (2000, pp. 607-608) argues that predominant sensitivity to scope presented in the review by Carson (1997) may primarily be a feature of applying internal tests, such that respondents are influenced by previously stated values and tries "to act in an internally consistent way by providing higher bids for larger magnitudes".[2]

Svedsäter (2000) concluded with insensitivity to scope in external scope tests. However, there exist CV studies on environmental goods with passive-use value that have rejected scope insensitivity in external tests. Hoevenagel (1996) concluded with sensitivity to scope (no perfect embedding) based on external scope tests of environmental projects in the Netherlands, and in a similar study from Norway, Magnussen (1992) reached the same conclusion. Giraud *et al.* (1999) reject scope insensitivity in external tests comparing WTP for two species protection plans. But, to their surprise, they did not find significant scope

[2] It was also the NOAA panel's view that the same survey respondent forces scale differentiation when faced with two goods in one survey (Arrow *et al.* 1993).

sensitivity in internal tests. They explain the finding by a "baseline bias" (a sequencing effect), that respondents based their WTP decision on the first good being valued.

The test of insensitivity to scope in this paper was based on a CV study of preserving endangered species in Norwegian forests. Different from most valuation studies of endangered species, this CV study focussed mainly "low-profile" species, i.e. insects and cryptogams, together with flowers and birds, rather than the "charismatic mega-fauna" (Brown and Shogren 1998). Most of the lower species arguably represent a complex environmental amenity with predominant passive-use values. The respondents to the CV survey were split into four sub-samples that valued different scopes in different sequences of endangered species preservation. Further, all four sub-samples were split according to elicitation format: Open ended *with* payment card versus open ended *without* payment card. The design of the study allowed for both external and internal scope tests.

The outline of the rest of the paper is as follows: Section 2 presents relevant theory for valuation of goods with different scope. In section 3 the data and scope test design are described. Section 5 contains the results, and these are discussed and concluded in the last two sections.

2. Theory and scope sensitivity hypotheses

Testing for scope effects over a goods composition it is only appropriate to operate with *nested* goods (Carson and Mitchell 1995). Assume a principal good/package, A, and define its subsets as b, c and d. A itself (equal to $b+c+d$) may be part of a larger package, say, the multi-

package **P**. Thus, the nesting relation is that *b*, *c* and *d* are subsets of *A*, and *A* is a subset of **P**. The nesting is categorical (not quantitative), whereby the more comprehensive good is composed of distinct subsets.

The sensitivity in CV of these three scope levels can be tested internally (within-subject) or externally (between-subject). Define valuation order within a sub-sample with numerical superscripts, and denote sub-sample with Roman numerical subscripts. Consider the following valuations in four sub-samples:

Sub-sample *I*	Sub-sample *II*	Sub-sample *III*	Sub-sample *IV*
A_I^1	c_{II}^1	b_{III}^1	$\mathbf{P}_{IV}^1$
	A_{II}^2	$(b+c)_{III}^2$	A_{IV}^2
		A_{III}^3	

External comparisons should primarily comprise only first-valued goods, to avoid the scope test to be biased by expected sequencing effects. However, in sub-samples *II* and *III* the sequences are "built-up" with increments up to *A*, such that a baseline/sequencing effect per se should not be expected to clutter the scope relationship. For the same reason internal scope tests between first-valued and second- or third-valued goods are "acceptable" when valuing increments (and decrements) in a nested goods' sequence (Carson and Mitchell 1995, Randall and Hoehn 1996).[3]

The external tests of insensitivity to scope (null hypothesis) based on the given valuations are:

[3] This type of internal scope test is also called *nested sequence testing*, while the external test represents *non-nested sequence testing* (Carson and Mitchell (1995).

$$\begin{aligned} &\mathrm{WTP}\left[c_{II}^{1}\right], \mathrm{WTP}\left[b_{III}^{1}\right] \leq \mathrm{WTP}\left[A_{I}^{1}\right] \leq \mathrm{WTP}\left[\mathbf{P}_{IV}^{1}\right] \\ &\mathrm{WTP}\left[c_{II}^{1}\right] \leq \mathrm{WTP}\left[(b+c)_{III}^{2}\right] \end{aligned} \quad (1)$$

The internal tests based on the given valuations are:

$$\begin{aligned} &\mathrm{WTP}\left[c_{II}^{1}\right] \leq \mathrm{WTP}\left[A_{II}^{2}\right] \\ &\mathrm{WTP}\left[b_{III}^{1}\right] \leq \mathrm{WTP}\left[(b+c)_{III}^{2}\right] \leq \mathrm{WTP}\left[A_{III}^{3}\right] \\ &\mathrm{WTP}\left[A_{IV}^{2}\right] \leq \mathrm{WTP}\left[\mathbf{P}_{IV}^{1}\right] \end{aligned} \quad (2)$$

Assuming local non-satiation (disregarding local "flat" utility), that is, the CV respondent obtains positive utility from the complement of the valued goods up to **P** (b^*, equal to c+d, c^*, equal to b+d, and A^*, equal to **P**-A), the weak inequality signs can be replaced by strong inequalities (Carson and Mitchell 1995).[4]

Assuming "well-behaved preferences" and that the good is normal, an incremental increase in the scope of the environmental good/package, should produce an increase in WTP (Boyle *et al.* 1994, Carson and Mitchell 1993). Even in the case where preferences are heterogeneous and part of the population are indifferent to any proposed (further) increase in the most comprehensive good (A or **P**) when "having" a subset (b or c, or A), it is deemed against common sense that a "large" increase in the composition of the good that is valued should not affect sample aggregate WTP (Boyle *et al.* 1994, 1998, Carson and Mitchell 1993, 1995, Giraud *et al.* 1999).

[4] Local satiation rules out globally maximising behaviour, which is a standard neo-classical assumption (Carson and Mitchell 1995, Varian 1992), but it does not conflict the basic axioms of a rational economic woman (complete, reflexive and transitive preferences).

Still, Rollins and Lyke (1998) hold that although preferences are strictly monotonic (non-satiated), *diminishing marginal value* implies that it becomes more difficult (statistically) to reject the null hypothesis of scope insensitivity with successive increases in quantities. They base the argument on an experiment involving valuation of different numbers of proposed parks (from 1 to 10) in Canada. Their results showed a greater difference in WTP between "the first proposed parks" compared to difference in WTP between "the last proposed parks". The implication is that it demands a larger sample to reject a null hypothesis of insensitivity to scope comparing WTP for the 9^{th} park versus the 10^{th} park than to reject scope insensitivity comparing WTP for the first versus the second park.

But, while the scope test in Rollins and Lyke (1998) is closer to test over numbers (although one park is not a perfect substitute for another), the scope tests in this paper are over the composition of a commodity being valued (Boyle *et al.* 1998). The subsets *b* and *c* may be regarded as too different to result in close substitutes (e.g., endangered animals versus endangered plants). However, one might imagine that the substitution effect, defined in terms of how a change in one good changes the marginal valuation of another good, can be more pronounced for the valuation in sequence, i.e., the internal scope relation (Hoehn and Loomis 1993, Giraud *et al.* 1999).

Income effects are often disregarded in explaining stated WTP in CV, with the argument that WTP for the good in question normally represents a relatively very low amount of household wealth or income (Mitchell and Carson 1989). However, discretionary budgets are much less than household wealth and income, since many kinds of expenditures are fixed in the short run. This may lead to *de facto* substitution between different types of goods, since one (hypothetical) purchase can strongly diminish or even eliminate the budgets that were

available after covering fixed expenditures. Thus, constrained optimisation may explain that some respondents will not pay for, e.g., *c* (or *A*) after holding *b* (Randall and Hoehn 1996).

The psychological explanations of observed insensitivity to scope predict very distinct outcomes for the CV method. According to the *good cause dump* and *purchase of moral satisfaction* hypotheses (Kahneman and Knetsch 1992) the null hypotheses of scope insensitivity will not be rejected in any cases. The conclusion of such an observation is theoretical invalidity of CV. On the contrary, the explanations by Green and Tunstall (1999) do not predict scope insensitivity on the whole and invalidation of CV. They argue that observation of insensitivity to scope is primarily due to the failure of asking WTP questions that don't fit the cognitive structure, or categorisation of the respondents. Their underlying argument is that preferences and WTP, like cognition, are held for categories, and that the (first) question about WTP has to hit the basic-category level. If the basic-category contains only **P**, *A* or *b*/*c*, then the external tests are not expected to provide valid results, since at least one of the goods' scope in the external comparison will probably be misunderstood. This would be a survey-design feature similar to what Mitchell and Carson (1989) term *amenity misspecification* (*symbolic*) bias. The internal tests will not be expected to show scope sensitivity if the first good does not hit the basic-category level.

3. The CV study

The data were collected from a survey of a representative sample of the Norwegian population.[5] The sample comprised a total of 1019 respondents, each taken to represent her household. The interviews were performed face-to-face in the respondents' home. The sample was split into 4 sub-samples that faced a different sequence of valuation of endangered species preservation (table 1). Further, each sub-sample was split according to elicitation format, either open-ended question of maximal WTP *with* the aid of a payment card (oe-pc), or open-ended question *without* any use of payment card (oe).

Table 1. Experimental design over nested environmental goods in four sub-samples (*S*).[a]

	Sub-sample *I* (*n*=260)	Sub-sample *II* (*n*=259)	Sub-sample *III* (*n*=255)	Sub-sample *IV* (*n*=245)
1st valuation	*A*	*c*	*b*	**P**[b]
2nd valuation		*A*	*b*+*c*	*A*
3rd valuation			*A*	
Nesting		$c \subset A$	$b \subset (b+c) \subset A$	$A \subset \mathbf{P}$

[a] *A* refers to preservation of *all* endangered forest species; *c* refers to endangered *cryptogams*, comprising fungi, lichen, and mosses; *b* refers to preservation of the white-backed woodpecker, which was the only bird in Norwegian forests denoted as *endangered* at the time of the survey. *b*+*c* is the sum/union of *b* and *c*; **P** refers to a multi-package of six environmental projects, mostly specific reductions of different types of pollution, in addition to *A*.

[b] Sub-sample *IV* initially valued "the three preferred (of the six) environmental projects", that did not necessarily include preserving endangered forest species (*A*).

All four sub-samples of the CV survey were presented the same reference condition for endangered species in Norwegian forests. Endangeredness was described as stemming mainly

[5] The survey was conducted by the Norwegian Gallup Institute (NGI) in March 1992 as part of a monthly *omnibus* dealing with several topics. The sample was produced using a two-stage stratified cluster sampling technique. This was based on municipalities as the unit, which first were stratified by part of the country. Then towns with more than 30,000 inhabitants were defined as own strata, while the other municipalities were classified according to structure of industry and centrality (designed by Statistics Norway). Household addresses were randomly selected for each sampling location, and the household member (15 and older) with the most recent birthday was interviewed. According to the Norwegian Gallup Institute this design should produce samples that approximate the adult civilian population. The Norwegian Gallup Institute also assisted in designing and pre-testing the questions for valuing endangered species.

from the decline in area covered with old-growth forest containing dead or dying trees, as a result of current forest management practices. A "worst-case baseline/scenario" was implied by the statement "if nothing is done, all these species can be expected to disappear". It was specified that most of the "several hundred" endangered species were *cryptogams* (i.e. fungi, lichen, and mosses) and insects. These species have no commercial use and are probably, in most cases, not observed during recreational activities. If individuals attach value to the preservation of such species, one might assume that *passive-use* values form a major component of total value. Other species mentioned in the survey were flowers and birds.

The target of the projects to be valued was "certain" survival, by means of more conscious forest management practices and preservation of larger forest areas. Sub-sample *I* valued only the "full scope" of endangered species preservation, that is, how much their household maximally was willing to pay per year to be sure that *all* endangered plants and animals in Norwegian forests (*A*) would be preserved. Sub-sample *II* initially valued the subset "preservation of endangered fungi, lichen and mosses in Norwegian forests", the cryptogams (*c*), before valuing the preservation of all endangered forest species (*A*). Sub-sample *III* was given some additional information about the white-backed woodpecker, and asked initially what was their household's maximal WTP to be sure that this bird (*b*) would be preserved.[6] Then sub-sample *III* valued the combined set of both the white-backed woodpecker and endangered cryptogams (*b*+*c*), and finally *A*. Sub-sample *IV* valued initially a more comprehensive package of six major environmental projects (**P**), including *A*, before valuing *A* only (table 1).[7]

[6] Respondents in all sub-samples were told that *ca* 10 birds in Norwegian forests are classified as *vulnerable*. The white-backed woodpecker was the only (non-migratory) bird classified as *threatened* in Norwegian forests.

The scenario in this survey can be described as complex, presenting endangered species in forests that were mostly unknown to the average Norwegian. It was also complex because it was difficult to delineate the amenity for the respondent, compared to an alternative approach focussing and valuing the *habitat* (Rollins and Lyke 1998). This also complicated the relation between the amenity to be valued (endangered species preservation) and the project to realise it (habitat preservation/adjustments). The valuation could also be complex merely due to the predominant existence- or passive-use value. Thus, turning the argument of Cummings *et al.* (1995): If CV can get the theoretically correct result with a set of the most complex public goods it should generally ”get it right” with public goods.[8]

The respondents were given some visual aid, in addition to the oral presentation of endangered species. A poster was also showed, made up of a large photo of an old-growth forest, surrounded with photos of endangered cryptogams, insects and the white-backed woodpecker. Some cards with main points concerning the scenario were also showed the respondents to facilitate their task. It was informed about the “residuals”/complements of both $A\text{-}c \approx$ endangered birds, flowers and insects ($b+d$), and $A\text{-}b \approx$ endangered cryptogams, flowers, and insects ($c+d$). It was assumed that the scenario together with cards and the poster with photos would help the respondent to, at least, distinguish between the categories *cryptogams* (several photos of fungi and lichen), *flowers* (one photo of an orchid), *birds* (a photo of the white-backed woodpecker), and *insects* (several photos).[9]

[7] The other five sub-packages of **P** involved 50% reductions of various pollutions. In the overall ranking of the six, *A* was ranked four (Veisten *et al.* 1993).

[8] The argument in Cummings *et al.* (1995, footnote 4, p. 261) was the following: “if CV [method] subjects cannot ’get it right’ with a private good, we believe that it is unlikely that they will get it right with a public good”. Carson *et al.* (2000) discuss why results based on CV of private goods cannot be readily transferred to a case involving (rationed) public goods, and *vice versa*, due to different incentive structures.

[9] Some respondents could possibly include the wolf and other endangered or vulnerable mammals, in the ‘preservation of *all* endangered species in Norwegian forests’ (*A*), even if these species are not within the group of endangered *forest* species.

The payment vehicle was presented as the following:

> "Suppose that the government introduced an environmental tax on wood products (lumber, paper and other wood-based products), to preserve *all* the plants and animals in danger of disappearing from Norwegian forests. Such an environmental tax would entail an increase in household expenditures."

Given that the respondents found it credible that such an environmental tax could be introduced, this represents a coercive payment mechanism (Carson *et al.* 2000, Sinden 1999). The elicitation format used was open-ended questions about maximal willingness to pay. For half of the sample of respondents the open-ended format was combined with the use of a payment card (oe-pc), having amounts from 0 up to US$1,840 (NOK 12,000) on a linear scale. There were no benchmarks indicating current payments (taxes) for the amenities in question or the costs of changes. For the other half of the sub-sample the open-ended format was used without additional tools (oe). It is argued that the use of a payment card may facilitate the respondents' task of putting a money value to the amenity in question, and hence reduce the regularly high amount of item non-responses and zero bids when asking about maximal WTP, but there is a possibility of a *range bias* (anchoring effect) related to a card with a given range of amounts (Mitchell and Carson 1989, Rowe *et al.* 1996).

In addition to the questions of maximal WTP to protect endangered species, the CV survey included follow-up questions about reasons to be willing to pay or reasons for stating zero WTP. Further, it included questions about respondents' attitude towards preservation of endangered species, their reaction towards a disappearance of a species from the forest they use, their recreational activity in forests, and their demographic characteristics (see Appendix). Some of these questions might serve as proxy variables for financial resources and tastes in a valuation function of WTP. The valuation function can be estimated as a regression model and used to test the external scope insensitivity hypothesis. It allows a

control for differences in personal characteristics between sub-samples. Further the valuation function serves as a test of theoretical validity by assessing whether WTP varies with a set of variables as predicted by economic theory and empirical findings (Arrow *et al.* 1993, Carson *et al.* 1994, Mitchell and Carson 1989).

Table 2. Relevant questions for a valuation function of WTP.

Type of variable	Question [Q#] [a]	Effect [b]
Elicitation of WTP	Open-ended question of maximal WTP [Q1.3]	
Financial resources	Household income	+
	Respondent or spouse unemployed	-
Tastes (use of forest)	# of trips to forest for different recreational activities during the last year [Q1.1]	+
Tastes (attitudes)	Negative reaction toward disappearance of species from the forest used for recreation [Q1.9]	+
	Level of concern for species preservation [Q1.2]	+
Tastes (knowledge)	Knowledge about a preservation plan and about forest management [Q1.11-12]	?
Tastes (demographic)	Urbanisation level	+
	Educational level	+
	Age	-
	Gender (male versus female)	-
	Political affiliation ("green" versus "non-green")	+

[a] Refers to question numbers for sub-sample *I* (given in the Appendix).

[b] *Effect* indicates the covariates' expected effect on WTP for preserving endangered species.

The basic economic variable in the valuation function is the financial resource (or available wealth) of the agents, which was proxied with their stated household income.[10] A dummy that equals one if either the respondent or the spouse was unemployed was also included. The agents' tastes for the environmental good were proxied by their use of the forest for recreation, their attitudes towards preserving species, and their demographic characteristics. The resource use for recreation was measured by a sum variable of the number of trips to forest areas for different recreational purposes.[11] The valuation function would also include demographic variables that have been commonly used as indicators of tastes: age, educational

[10] Measuring financial resource using respondents' stated household income, will often incur problems of missing data and measurement error (Greene 1997, Carson *et al.* 1994, Mitchell and Carson 1989).

[11] This included hiking, cross-country skiing, hunting, picking berries and mushrooms.

level and urbanisation level (Carson *et al.* 1994, Mitchell and Carson 1989). The hypothesised effect from the covariates is based on (economic) theoretical expectations, as for income and unemployment, and on (psychological) expectations of correlation between expressed tastes for the environment and WTP. Earlier findings from CV studies of environmental goods show mainly decreasing WTP by age and increasing WTP by education and urbanisation (Carson *et al.* 1994, Giraud *et al.* 1999, Loomis *et al.* 1993, Veisten *et al.* 1993), although without providing theoretical explanations (table 2).

The external and internal scope hypotheses were tested based both on a comparison of sample means (*t*-test) and non-parametric distributions of WTP (Mann-Whitney test). From Tobit models with covariates likelihood-ratio tests of equality were performed for the external tests (Greene 1997). The Tobit model allows for indifference (no WTP for increasing the baseline or avoiding the extinction) and was chosen as econometric model of WTP, as a relatively large part of respondents stated 0 WTP. Since the positive amounts of hypothetically stated WTP had a skewed distribution with a long right tail, a model based on log-transformation bringing the distribution of WTP closer to normal was applied (Schulze *et al.* 1996, Giraud *et al.* 1999). All statistical analysis was done with the statistical package SAS 6.12 (SAS 1997).

4. Results

4.1. Descriptive statistics

Summary statistics are split on elicitation technique: open-ended (oe) questions without any aid and open ended with the aid of a payment card (oe-pc). A quick look at median WTP in the "raw" (non-cleaned) samples indicates that the differently sized goods mostly line-up as one would expect them to do. Preserving all endangered species in Norwegian forests (*A*) gets higher median values than preserving subsets of this comprehensive good (*b*, *c*, *b*+*c*), and *A* gets lower median values than the environmental multi-package (**P**) of which *A* was one of six components. A possible indication of oddity relates to median WTP for preserving endangered cryptogams (*c*) in sub-sample *II*. The amount is for *c* in *II* is higher than for preserving the endangered bird *plus* the cryptogams (*b*+*c*) in sub-sample *III* (tables 3a and 3b).

There were less item non-responses to the WTP questions, in the raw data, when respondents stated maximum WTP with the aid of the payment card (oe-pc), relative to the open-ended elicitation format without payment card (oe). Item non-responses reduce the sizes of the sub-samples by about 9-15% in the oe case (table 3a), and by about 1.5-5.5% in the oe-pc case (table 3b). The oe-pc format also gave less zero responses. Further, more respondents stated WTP above 2% of their household income (for preserving *A*) with the oe-pc format than with the oe format. Overall, elicitation of maximal WTP using oe-pc has generally produced higher estimates than when using oe (tables 3a and 3b).

Table 3a. WTP sample statistics – open-ended questions without payment card (oe**).** [a]

	Sub-sample / Goods [b]							
	I	*II*		*III*			*IV*	
Statistic	*A*	*c*	*A*	*b*	*b+c*	*A*	**P**	*A*
	Raw data							
n (raw)	135	127		131			116	
Median	$38.25	$20.65	$76.50	$15.30	$15.30	$30.60	$76.50	$30.60
Maximum	$3,672	$1,836	$1,836	$459	$1,530	$4,590	$3,060	$5,355
0	35.0%	38.9%	34.5%	41.7%	36.9%	29.9%	38.1%	37.5%
n (non-miss.)	123	108	110	115	114	114	105	104
Item nonr.	8.9%	15.0%	13.4%	12.2%	13.0%	13.0%	9.5%	10.3%
>2% inc.	2.2%	6.3%		1.5%			2.6%	
Cleaning [c]	14.8%	12.6%		12.2%			14.7%	
	Cleaned data [c]							
n (clean)	109	106		108			96	
Mean	**$104.27**	**$107.66**	**$130.00**	**$40.62**	**$55.38**	**$92.93**	**$234.69**	**$79.08**
SE	196.34	265.75	229.99	81.49	95.24	163.10	482.27	133.36
Median	**$42.07**	**$30.60**	**$76.50**	**$15.30**	**$15.30**	**$30.60**	**$76.50**	**$30.60**
Maximum	$1,530	$1,836	$1,836	$459	$459	$918	$3,060	$918
0	33.0%	33.2%	29.3%	38.6%	32.6%	25.2%	39.1%	38.4%
n (non-miss.)	100	88	90	96	95	95	87	86
Item nonr.	8.3%	17.0%	15.1%	11.1%	12.0%	12.0%	9.4%	10.4%
Equal WTP [d]		27.4%		19.4%			18.8%	

[a] Estimates given in USD ($). USD 1 ≈ NOK 6.5359 – the average exchange rate through March 1992. (NOK, Norwegian *Kroner*, was applied in the CV survey.) All statistics related to WTP and zero percentages are calculated for non-missing data.

[b] The goods and their relationship are described in table 1.

[c] Observations deleted if WTP>2% of income, income missing, protest answers, or preference inconsistencies.

[d] Equal WTP are percentages of sub-samples stating equal positive amounts for all goods valued. In sub-sample *III* the percentages of equal WTP were higher for two-good comparisons, up to 27.8% stated equal WTP for *b+c* and for *A*.

Table 3b. WTP sample statistics – open-ended questions with payment card (oe-pc**).** [a]

	Sub-sample / Goods [b]							
	I	*II*		*III*			*IV*	
Statistic	*A*	*c*	*A*	*b*	*A*	*C*	*A*	*b*
	Raw data							
n (raw)	125	132		124			129	
Median	$76.50	$38.25	$76.50	$15.30	$30.60	$76.50	$153.00	$45.90
Maximum	$4,590	$1,530	$13,770	$459	$612	$3,060	$12,240	$12,240
0	26.0%	32.8%	28.4%	40.8%	41.2%	33.9%	36.3%	34.5%
n (non-miss.)	123	128	127	120	119	118	127	122
Item nonr.	1.7%	3.0%	3.8%	3.2%	4.0%	4.8%	1.6%	5.4%
>2% inc.	7.2%	9.1%		4.8%			9.3%	
Cleaning [c]	24.0%	16.7%		10.5%			16.3%	
	Cleaned data [c]							
n (clean)	93	104		101			103	
Mean	**$160.57**	**$104.22**	**$152.61**	**$77.43**	**$86.46**	**$149.54**	**$352.50**	**$132.60**
SE	203.10	162.73	214.24	130.19	140.67	214.22	521.15	237.01
Median	**$76.50**	**$38.25**	**$76.50**	**$15.30**	**$30.60**	**$76.50**	**$114.75**	**$45.90**
Maximum	$918	$918	$918	$459	$612	$918	$1,836	$1,377
0	25.0%	35.6%	29.3%	40.2%	41.7%	32.6%	38.3%	35.0%
n (non-miss.)	94	101	99	97	96	95	102	97
Item nonr.	1.1%	2.9%	4.8%	4.0%	5.0%	5.9%	1.0%	5.8%
Equal WTP [d]		34.6%		17.8%			13.6%	

[a] Estimates given in USD ($). USD 1 ≈ NOK 6.5359 – the average exchange rate through March 1992. (NOK, Norwegian *Kroner*, was applied in the CV survey.) All statistics related to WTP and zero percentages are calculated for non-missing data.

[b] The goods and their relationship are described in table 1.

[c] Observations deleted if WTP>2% of income, income missing, protest answers, or preference inconsistencies.

[d] Equal WTP are percentages of sub-samples stating equal positive amounts for all goods valued. In sub-sample *III* the percentages of equal WTP were higher for two-good comparisons, up to 24.3% stated equal WTP for *b*+*c* and for *A*, and for *b* and *b*+*c*.

Following the standard practice in CV studies, the raw data were "cleaned" before reporting mean values (Mitchell and Carson 1989, Carson *et al.* 1994, Rowe *et al.* 1996). The cleaning procedure deleted those who stated WTP above 2% of their household's income, those not reporting household income, and those who indicated contradictory preferences.[12] This cleaning procedure indirectly eliminated outliers with stated WTP for preserving endangered species above the range of the payment card, USD 1,836 (NOK 12,000). Only the maximum value for the multi-package **P** remained higher than USD 1,836. The cleaning reduced the sub-samples' size by 10.5% to 24%.

Looking at the mean values in the cleaned sub-samples, the oe-pc values of mean WTP had exactly the same relationship as the median values in the raw samples. Thus, only the relation between *c* and *b*+*c* remained odd (table 3b). Turning to the oe estimates there were more odd results for the external scope relations: Mean stated WTP for *c* in Sub-sample *II* was higher than the means for *A* in all other sub-samples (table 3a).

The percentage of zero WTP in non-missing cleaned data also varied similarly with the scope of the good, with more zeros for the smaller goods than for the larger; and also in that case more zeros for *b* than for *c* (tables 3a and 3b). Another interesting aspect about apparent insensitivity to scope was the registered percentage of equal stated positive WTP for different goods in the sub-samples. This percentage was highest in sub-sample *II*, with 27% in the oe

[12] 5.4% of the total "raw" sample (*n*=1019) stated WTP above 2% of their household's annual gross income, and 9.8% did not report income. Protest zeros amounted to 4%, and included those who did not accept the payment vehicle, the assumption of property rights, the probability-of-provision, or the monetary valuation of species as such (Q1.4), and, at the same time stated that they were "very engaged" by the issue of preserving endangered plants and animals or "working actively with preservation of animals and plants" (Q1.2). Those who answered "don't regard extinction of species as a problem" (Q1.2) and at the same time stated positive WTP to preserve endangered species were defined as inconsistent answers. This group counted 0.9% of the raw sample (see Appendix and Veisten *et al.* 1993).

case and nearly 35% in the oe-pc case. In sub-samples *III* (equal WTP for *b*, *b*+*c* and *A*) and *IV* this percentage oscillated between 14% and 19% (tables 3a and 3b).

4.2. Basic scope tests

For the external scope test the simple *t* test gave rejection of insensitivity to scope in three of four tests in the payment card (oe-pc) case. In the oe and pooled (oe – oe-pc) cases, the null hypothesis was not rejected when testing equality of WTP for *c* in sub-sample *II* and *A* in sub-sample *I*. The Mann-Whitney (MW) rank sum test, on the contrary, gave rejection of insensitivity to scope comparing *c* and *A* in the pooled model, but did not reject equality between *A* in sub-sample *I* and the multi-package **P** in sub-sample *IV*. Overall, the MW test gave slightly fewer rejections of the null hypothesis than the *t* test. Since WTP for *c* in sub-sample *II* was clearly higher than WTP for (*b*+*c*) in sub-sample *III*, this was included as non-rejection of external scope insensitivity (table 4).

Table 4. External scope tests (one-sided). [a]

	Pooled oe – oe-pc		oe		oe-pc		Overall test result
Null hypothesis	*t* test [b]	MW [c]	*t* test [b]	MW [c]	*t* test [b]	MW [c]	
$\hat{WTP}[A_I^1] \geq \hat{WTP}[c_{II}^1]$	1.19	2.09**	0.10	0.76	2.14**	2.32**	Mixed
$\hat{WTP}[A_I^1] \geq \hat{WTP}[b_{III}^1]$	4.37***	4.70***	2.94***	3.09***	3.37***	3.62***	Rejected
$\hat{WTP}[\mathbf{P}_{IV}^1] \geq \hat{WTP}[A_I^1]$	4.25***	1.19	2.48***	0.94	3.31***	0.73	Mixed
$\hat{WTP}[(b+c)_{III}^2] \geq \hat{WTP}[c_{II}^1]$							Not rejected [d]

[a] Tests are based on cleaned data sets. Sample sizes are given in tables 3a and 3b.
[b] The *t* test gives the student T value assuming equal variances.
[c] MW is the Mann-Whitney rank sum test, and the presented MW test statistic is a normal-approximated T value.
[d] No statistics are given since both mean WTP and the ranked distribution were lower for (*b*+*c*) than for *c*, such that $\hat{WTP}[(b+c)_{III}^2] \geq \hat{WTP}[c_{II}^1]$ is obviously not rejected.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

For the internal scope test, both the *t* test and the MW test rejected the null hypotheses of insensitivity to scope in three out of five tests, independently of elicitation format. Equality between WTP for *A* and WTP for *c* in sub-sample *II* is only rejected using the *t*-test for the oe-pc format, and the MW test for the pooled (oe – oe-pc). Equality between WTP for *b* and WTP for *b*+*c* in sub-sample *III* is only rejected in one case, and only at the 10% level (table 5). In the case of internal scope tests, the MW test gave slightly more rejections of the null hypothesis than the *t* test.

For both internal and external scope tests, given that the WTP distributions were not normal and also had a spike at 0 WTP, there is reason to put relatively more weight on the non-parametric MW test than the *t* test (Boyle *et al.* 1994).

Table 5. Internal scope tests (one-sided). [a]

Null hypothesis	Pooled oe – oe-pc		oe		oe-pc		Overall test result
	t test [b]	MW [c]	*t* test [b]	MW [c]	*t* test [b]	MW [c]	
$\hat{WTP}[A^2_{II}] \geq \hat{WTP}[c^1_{II}]$	1.60*	2.19**	0.60	1.57*	1.80*	1.56*	Mixed
$\hat{WTP}[A^3_{III}] \geq \hat{WTP}[b^1_{III}]$	3.89***	3.95***	2.81***	3.35***	2.83***	2.47***	Rejected
$\hat{WTP}[(b+c)^2_{III}] \geq \hat{WTP}[b^1_{III}]$	1.01	1.10	1.15	1.45*	0.47	0.23	Mixed
$\hat{WTP}[A^3_{III}] \geq \hat{WTP}[(b+c)^2_{III}]$	3.06***	2.89***	1.94*	1.98**	2.41***	2.23**	Rejected
$\hat{WTP}[\mathbf{P}^1_{IV}] \geq \hat{WTP}[A^2_{IV}]$	4.77***	2.63***	2.89***	1.67**	3.80***	2.08**	Rejected

[a] Tests are based on cleaned data sets. Sample sizes are given in tables 3a and 3b.
[b] The *t* test gives the student T value assuming equal variances.
[c] MW is the Mann-Whitney rank sum test, and the presented MW test statistic is a normal-approximated T value.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

5.3. Retesting external scope applying a Tobit model of WTP

Willingness to pay (WTP) for the different scopes of endangered species preservation, plus the multi-package **P**, were regressed on a set of demographic and attitudinal variables

applying the Tobit model. WTP was log transformed, since it brought the distribution of positive WTP closer to normality, and it gave a better fit than using untransformed WTP.[13] The choice of a "best explanatory model" was based on a total (cleaned) sample regression for *A*, using a backward selection on 15% level among the plausible covariates. Knowledge, gender and political voting did not show up with significant signs (table 6).

Table 6. Mean values of variables used in Tobit regression analysis. [a]

Variable	Description	Sub-samples			
		I (*n*=202)	*II* (*n*=210)	*III* (*n*=209)	*IV* (*n*=199)
LNINCOME	Log of gross household income [b]	42.7	42.8	41.3	43.1
LNAGE	Log of age [b]	41.9	42.2	42.6	42.4
LNTRIPS	Log of number of visits to forests (category midpoints) [b]	30.1	30.7	31.2	32.0
EDUC	Education, 3 levels (1-3)	1.92	1.94	1.91	1.89
URBAN	Urbanisation, 6 levels from rural (1) to "bigger city" (6)	2.77	2.73	2.71	2.68
ENGPRE	Engagement in preserving endangered species, 4 levels, from "no problem" (0) to "working actively" (3)	1.36	1.33	1.36	1.31
REDU	Disappearance of species from the forest used for recreation reduce satisfaction = 1, 0 otherwise	0.71	0.76	0.70	0.74
UNEMP	Respondent or spouse unemployed = 1, 0 otherwise	0.19	0.18	0.17	0.18
CARD	Payment card used to elicit max WTP =1, 0 otherwise	0.50	0.50	0.50	0.50
CLEAN	Deleted observations with WTP>2% of income, or income missing	22.3%	18.9%	18.0%	18.8%

[a] Pooled oe – oe-pc (cleaned).

[b] Mean values from untransformed numbers, income in $1,000.

Then the Tobit model was estimated for all the goods, both split on elicitation format and pooled. For convenience only the pooled results are presented in tables (8a and 8b). Only one of the covariates had significant expected sign in the regression of WTP for any of the goods: (log of) household income (LNINCOME). Stated engagement in preserving endangered species (ENGPRE) mostly had a significant positive sign, and (log of) age (LNAGE) mostly has a significantly negative sign. The rest of the covariates prevailingly came out with plausible signs, but only some had significant signs in more than one case (EDUC, URBAN,

[13] Although normality of the log-transformed positive WTP for any good in any sub-sample could mostly be rejected on the 5% level, the Kolmogorov statistic unanimously came closer to unity with the log-transformation.

UNEMP). (Log of) number of visits to the forest (LNTRIPS) had an implausible negative sign (although not significant) for the WTP of the goods evaluated by sub-sample *III*. The variable indicating respondent or spouse being unemployed (UNEMP) had an implausible positive sign (although not significant) for the WTP of the goods evaluated by sub-sample *II*. The use of a payment card only has a significantly positive sign for WTP for *A* by sub-sample *I* (tables 7a and 7b).

Table 7a. Estimated Tobit regression models – first-valued goods. [a]

Covariates [b]	Good / Sub-sample [c]			
	$\ln \hat{WTP}[A_I^1]$	$\ln \hat{WTP}[c_{II}^1]$	$\ln \hat{WTP}[b_{III}^1]$	$\ln \hat{WTP}[\mathbf{P}_{IV}^1]$
Intercept	-15.8 (6.92)**	-16.7 (6.79)**	-13.1 (7.82)*	-2.63 (8.56)
LNINCOME	1.38 (0.51)***	1.27 (0.51)**	1.44 (0.54)***	1.26 (0.63)**
LNTRIPS	0.62 (0.27)**	0.24 (0.30)	-0.09 (0.27)	0.53 (0.37)
LNAGE	-1.29 (0.68)*	-0.62 (0.82)	-1.75 (0.84)**	-4.74 (0.96)***
EDUC	0.84 (0.44)*	0.46 (0.51)	0.70 (0.52)	1.19 (0.62)*
URBAN	0.32 (0.18)*	0.25 (0.23)	-0.12 (0.22)	0.50 (0.26)*
ENGPRE	1.31 (0.49)***	2.74 (0.61)***	2.03 (0.58)***	1.59 (0.64)**
REDU	1.40 (0.63)**	0.57 (0.81)	0.98 (0.73)	0.76 (0.90)
UNEMP	-1.58 (0.74)**	0.39 (0.87)	-0.76 (0.87)	-2.18 (1.01)**
CARD	1.33 (0.56)**	-0.08 (0.66)	0.36 (0.64)	-0.15 (0.76)
Normal Scale Parameter	3.66 (0.24)	4.13 (0.30)	4.07 (0.30)	4.75 (0.35)
Log likelihood	-420.1	-398.8	-391.5	-398.9
McFadden pseudo-R^2	6.49%	6.20%	4.97%	8.00%
n (non-missing)	190	186	192	187

[a] Pooled oe – oe-pc (cleaned).
[b] The covariates are explained in tables 2 and 8.
[c] Standard errors in parentheses.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

Table 7b. Estimated Tobit regression models – sequence-valued goods. [a]

Covariates [b]	Good / Sub-sample [c]			
	$\ln \hat{WTP}\left[A_{II}^{2}\right]$	$\ln \hat{WTP}\left[(b+c)_{III}^{2}\right]$	$\ln \hat{WTP}\left[A_{III}^{3}\right]$	$\ln \hat{WTP}\left[A_{IV}^{2}\right]$
Intercept	-12.5 (6.19)**	-15.1 (7.79)*	-13.1 (6.54)*	1.15 (7.11)
LNINCOME	1.13 (0.47)**	1.62 (0.54)***	1.39 (0.45)***	0.79 (0.52)**
LNTRIPS	0.60 (0.28)**	-0.11 (0.27)	-0.16 (0.23)	0.64 (0.31)
LNAGE	-1.06 (0.76)	-1.49 (0.84)*	-1.31 (0.71)*	-4.15 (0.80)***
EDUC	0.36 (0.46)	0.69 (0.51)	1.15 (0.43)***	0.63 (0.52)
URBAN	0.02 (0.21)	-0.07 (0.22)	-0.07 (0.19)	0.47 (0.21)**
ENGPRE	2.58 (0.56)***	1.79 (0.58)***	2.10 (0.49)***	1.79 (0.53)***
REDU	0.23 (0.75)	0.78 (0.72)	0.94 (0.61)	1.12 (0.76)
UNEMP	0.28 (0.80)	-0.61 (0.87)	-0.67 (0.73)	-2.34 (0.86)***
CARD	-0.03 (0.60)	-0.40 (0.63)	-0.16 (0.54)	-0.03 (0.63)
Normal Scale Parameter	3.84 (0.26)	4.07 (0.29)	3.50 (0.23)	3.91 (0.29)
Log likelihood	-412.8	-398.6	-411.4	-372.9
McFadden pseudo-R^2	6.63%	4.66%	6.88%	8.95%
# of observations	186	190	189	181

[a] Pooled oe – oe-pc (cleaned).
[b] The covariates are explained in tables 2 and 8.
[c] Standard errors in parentheses.
* $p < 0.10$
** $p < 0.05$
*** $p < 0.01$

Regarding goodness-of-fit measures, the pseudo R-square value indicated that the model for the "higher order" goods valued by sub-sample *IV*, *A* and **P**, showed relatively better fit, although the R-square values in all cases were generally low. The models for the "lower order" goods valued by sub-sample *III* (*b* and *b*+*c*) showed the lowest goodness-of-fit measures. Regarding significant signs of the covariates, the chosen econometric model shows best fit for WTP for *A* by sub-sample *I*. All covariates, chosen from the total cleaned sample estimation of WTP for *A*, had significant signs for sub-sample *I*'s WTP for *A* (tables 7a and 7b).

Based on log-likelihood values from the Tobit models, a likelihood ratio test was conducted for the external scope comparison (except the obvious case with non-rejected scope insensitivity, $\hat{WTP}\left[(b+c)_{III}^{2}\right] \geq \hat{WTP}\left[c_{II}^{1}\right]$). The likelihood ratio test value (q) is:

$$q = -2\left[\log L_{\text{pooled sub-samples}} - \left(\log L_{\text{x}} + \log L_{\text{y}}\right)\right] \sim \chi^{2}\left(d.f.\right) \qquad (4)$$

where logL_x and logL_y refer to the log likelihood values of the Tobit model for WTP for any of the goods from individual sub-samples, and log$L_{\text{pooled sub-samples}}$ is the log likelihood value based on a pooled sample of the two sub-samples.

Table 8. Likelihood ratio test (external scope) based on log-likelihood values from Tobit models.[a]

Null hypothesis	LRT [b]			Overall test result
	Pooled oe – oe-pc	oe	oe-pc	
$\widehat{\text{WTP}}[A_I^1] \geq \widehat{\text{WTP}}[c_{II}^1]$	18.6**	6.1	19.0**	Mixed
$\widehat{\text{WTP}}[A_I^1] \geq \widehat{\text{WTP}}[b_{III}^1]$	23.7***	13.7	22.5***	Mixed
$\widehat{\text{WTP}}[\mathbf{P}_{IV}^1] \geq \widehat{\text{WTP}}[A_I^1]$	21.2**	11.5	21.8***	Mixed

[a] Based on cleaned data sets. Sample sizes given in tables 10a and 10b.

[b] LRT is the likelihood ratio test, and the test value ($\hat{q}$) is compared to a chi-square distribution with K degrees of freedom (*d.f.*), where K is the sum of coefficients in the individual regressions, including the intercept, minus the number of coefficients in the pooled sub-samples regression, i.e., K=10. The critical value of χ^2 with 10 *d.f.* is ~16.0 at the 0.10 level, ~18.3 at the 0.05 level, and ~23.2 at the 0.01 level. For the elicitation-split sub-samples (oe and oe-pc) K=9, and the critical values are, respectively, ~14.7, ~16.9 and ~21.7.

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

Comparing the tests value ($\hat{q}$) to the chi-square distribution, with K=10 *d.f.* in the pooled oe – oe-pc case, leads to a rejection of the null hypothesis of insensitivity to external scope in two of the three cases, not for $\widehat{\text{WTP}}[A_I^1] \geq \widehat{\text{WTP}}[c_{II}^1]$. In the oe-pc case the three null hypotheses are rejected, while in the oe case no null hypotheses are rejected (table 8). The overall result from the likelihood ratio test is consistent with the combined result of the *t* test and the Mann-Whitney test.[14]

[14] The *t* tests and MW tests and the likelihood ratio tests were also run on the full (raw/non-cleaned) sample, pooled oe – oe-pc, leading to nearly identical test results. There were slightly stronger rejections of the null hypotheses in the external test $\widehat{\text{WTP}}[\mathbf{P}_{IV}^1] \geq \widehat{\text{WTP}}[A_I^1]$ and in the internal test $\widehat{\text{WTP}}[A_{II}^2] \geq \widehat{\text{WTP}}[c_{II}^1]$.

5. Discussion

The presented scope tests were based on a large-scale CV study done in-person by professional interviewers. Following up a study that focussed preservation of habitat and old coniferous forests (Hoen and Winther 1993), this study endeavoured a direct valuation of the lower species that make up the paramount share of endangered species in Norwegian forests. The scope effect results contribute to the assessment about the CV method's applicability to less known and complex elements of environmental amenities.

The preservation of *all* endangered species in Norwegian forests (*A*) was mostly found to be valued higher than preserving one subgroup or a single species. But, the difference between *A* and the subgroup of endangered fungi, lichens, and mosses (*c*) was not significant in the tests based on the halves of the sub-samples responding to open ended questions *without* payment card (oe).

The first valuation of *c* in sub-sample *II* seems to have resulted in an exaggerated WTP. In sub-sample *III* the addition of endangered cryptogams to the endangered bird, *b*, mostly did not increase significantly the WTP relatively to WTP for *b*. The preservation of endangered cryptogams either did not constitute a comprehensible initial task for economic valuation or simply was a too fuzzy environmental good for any CV task. Another explanation is that asking WTP for *c* did not hit the basic category level that most people hold for endangered-species preservation (Green and Tunstall 1999, p. 213). The basic category level would rather be the whole (*A*) or maybe the habitat (natural or virgin forests), such that asking WTP for *c* first would be meaningless for many respondents.

The bottom-up approaches in both sub-sample *II* and *III* may have been cognitively problematic if the respondents felt it strange/incorrect to initially state values for the preservation of a part of the totality *A*. Even if the single bird species was easier to grasp and delineate than cryptograms, one may still question if the initial valuation of *b* also led to invalid results in sub-sample *III*. The low goodness-of-fit values in the Tobit model could to some extent support this suspicion. Yet, the scope relation in sub-sample *III* turned out right, comparing mean WTP values, although the difference between WTP for *b* and for (*b*+*c*) was not significant.

Boyle *et al.* (1994) ask if their observed insensitivity to scope is a "true reflection of preferences", either that respondents' utility functions may contain flat spots because of satiation or CV is "incapable of measuring the desired incremental value". Flat spots or diminishing marginal utility were not easily assessed with the present data, since it compared compositions of endangered species preservation. But, differently from Boyle *et al.* most scope comparisons, at least with the payment-card application (oe-pc), show sensitivity to the scope of *A* versus its sub-components. Only for sub-sample *II*'s stated WTP for *c* it is strongly indicated that it does not truly reflect preferences for this sub-group of endangered forest species.

The results from the present study do not show such a clear discrepancy between external and internal tests as in, e.g., Giraud *et al.* (1999). There is one possibly important exception though. The non-parametric test rejected equality in the internal test of *A* versus **P** (oe-pc and pooled) but not in the external test. Following the explanation by Svedsäter (2000), the respondents in sub-sample *IV* seemingly pressed down their WTP for *A*, to make it consistent with their stated WTP for the multi-package **P**. One would expect clearer scope sensitivity

between *A* and **P** than between *A* and its sub-components, because the difference in scope between *A* and **P** was larger (Hoevenagel 1996, Magnussen 1992, McFadden and Leonard 1993). However, the inclusiveness of **P** may have constituted a counteracting effect to the scope difference. Due to the top-down approach the respondents in sub-sample *IV* faced initially a larger change in provision of public goods than the other sub-samples, and even if *A* was focussed afterwards, they should be expected to keep in mind the other sub-packages of **P**. The combined counteracting effect could have been that the budget constraint (income) effect has curbed WTP for **P** and that the WTP for *A* was curbed due the consideration of its complement (**P**-*A*). Loomis *et al.* (1993) found that apparent insensitivity to scope was more pronounced on the most inclusive levels of habitat protection in Australia. The difference in WTP between the largest and middle habitat proposal was less than the WTP difference between the smallest and the middle, although the scope difference was smallest between the small habitat proposal and the middle.

A considerable part of the respondents stated equal WTP for differently scoped goods in a sequence, even between *A* and **P** (14% to 19%, for oe-pc and oe format, respectively). It has been indicated that stating equal WTP for *c* and *A* in sub-sample *II* may be due to an upward amenity misspecification bias on WTP for *c*. But, it is not very likely that exactly equal WTP for *b* and *A* in sub-sample *III* or **P** and *A* in sub-sample can be due neither to cognitive problems nor to income effects / diminishing marginal values. The moral satisfaction purchase of good causes (Kahneman and Knetsch 1992), to which the scope of the amenity is approximately irrelevant, may be the adequate explanation for the valuation behaviour of these respondents.

The amenities valued in the present study were generally less familiar than those valued by, e.g., Svedsäter (2000) who included charismatic mega-fauna protection (elephant, rhino, tiger, koala, chimpanzee), rain forest protection, reduced global warming, and improved air quality. Yet, in this study the complex scenario was presented within a frame relevant for national policy. In Svedsäter only air quality was related to "the living area" of the respondent, and thus within a frame where the respondent would perceive that her answer would have some effect upon the policy maker that evaluates the proposal.[15] Valuing national projects for biodiversity preservation meet the criterion for a *consequential* survey (Carson *et al.* 2000), in contrast to the valuation of biodiversity in other countries without any clear clarification and assurance of how the respondent's values and money would lead to the realisation of the proposal.

This study presented a baseline extinction risk telling that "[if] we are not willing to spend money to cover the extra costs of … restrictions on harvesting, the endangered species in Norwegian forests might disappear". The result of the proposed preservation measures was cast as "certain" survival. This may be viewed as unrealistic, and ideally one would have included probabilities of survival (extinction) for different endangered species both with *no* preservation projects, and probabilities of survival *with* the projects. Primarily since ecologists/biologists could not provide such probabilities it was chosen to present the valuation choice between *extinction* and *certain survival*.

[15] While this study compared WTP for different scopes of environmental protection, Svedsäter (2000) compared WTP for *improvement* and *major improvement*, a verbal cue that is clearly inadequate to establishing a relevant scope relation. The apparent insensitivity to scope in Svedsäter's study, as in earlier well-cited studies concluding with scope insensitivity, can readily be explained by a flawed survey design that has resulted in *amenity misspecification* bias (Carson *et al.* 1997, 2001, Carson and Mitchell 1993, 1995, Hanemann 1994).

The elicitation procedure influenced on stated WTP and the statistical test of scope insensitivity. The results were more in accordance with expectations from economic theory for the half of the sub-samples stating WTP with the aid of a payment card (oe-pc).[16] With the aid of the payment card fewer respondents refused to answer and fewer stated zero. Both effects would contribute to strengthen the ability to reject scope insensitivity (increasing the observation points on the positive WTP distribution), as long as these respondents would not state equal WTP for all goods. The difficulty of responding to an open-ended question about maximum WTP, without any further aid to facilitate the task, may be a crucial element to obtain valid CV results (Hanemann 1994, Ready *et al.* 2001).[17] That WTP for *c* with oe format is close to the WTP for *c* with oe-pc format should not be taken as a sign of resilience, but rather providing additional evidence that the oe format, due to the extra cognitive challenge of "stating a price" without any aid, has "blown up" the upward amenity misspecification bias for *c*.

This research may be regarded as "probing the limits of where contingent valuation works" (Boyle *et al.* 1998) due to the complex and unfamiliar goods that were valued. Even if insensitivity to scope is rejected in most cases, the study does not provide any specific answer to the challenge of exhibiting a complex environmental amenity for valuation. Another issue that this study does not intend to answer is about how "large" a scope difference should be (Smith and Osborne 1996). While it is most unlikely and without any theoretical basis to expect WTP to increase linearly with number of parks, Km^2 or species, one might still ask if

[16] The elicitation-split sub-samples differed in terms of socio-economic characteristics that were correlated with WTP; both household income, age, and the level of education. Still, it is deemed unlikely that this should affect overall WTP difference between oe and oe-pc and difference in scope test results.

[17] Both stating equal positive WTP, zero WTP to all valuation questions or simply not answering the WTP questions (item non-response) contribute to distort the ability to reject insensitivity to scope. The indifferent group state legitimate answers according to economic theory (Kriström 1997), but due to the difficulty of stating a value (without any aid) some zeros may not represent true preferences.

rejecting the null hypothesis of insensitivity to scope is "enough" to conclude that CV provides the correct estimates in relation to scope or size.

6. Conclusions

The results of this study do indicate that some respondents may be truly insensitive to the scope of environmental amenities valued by CV. However, the behaviour of this minority of respondents did not have any dominating effect on the outcome of the scope tests. The effect of elicitation format had a more decisive effect. Use of an open-ended elicitation format without any aid to state values (oe) increased the number of respondents not answering the WTP question or stating a possibly untrue zero WTP. Insensitivity to scope is rejected in fewer tests with the halves of the sub-samples responding to the oe-format. On the contrary, for the other halves stating WTP with the aid of a payment card (oe-pc), insensitivity to scope was rejected in most cases.

The results did not demonstrate any clear difference between external and internal scope tests. In the case where an internal test gave stronger rejection than in the external, an income effect corroborated by a top-down approach with implicitly larger choice set can explain this alternatively to that respondents have tried to act in an internally consistent manner (with less regard to true preferences). It is an important result that insensitivity to scope is rejected in all clean external tests including only first-valued goods (except in the oe case).

Cognitive psychological issues related to the presentation of the scenario and describing the good remain the major factor of explaining apparent anomalies. Those tests that did not reject

insensitivity to scope all included WTP for the preservation of endangered cryptogams. Both the extra obscurity of this sub-group of endangered forest species and the probable misfit to the cognitive structure of respondents may have caused an amenity misspecification bias. The obvious complexity of this sub-group (compared to a single endangered bird species), and that the totality of endangered forest species is arguably closest to being the basic category level, enabled the singling out of this decisive bias. While other psychological and economic effects may have reduced the strength of the scope test, amenity misspecification rather crushed the whole thing resulting in apparent insensitivity to scope.

Although the communication of the specific change in the provision of an environmental amenity seems to be the crucial factor for obtaining valid WTP estimates in CV, other elements should not be passed over. Although the validity assessment is primarily assessed through (differences in) mean and median estimates and the overall WTP distribution, it should be of interest to investigate further how different respondents evaluate the task of stating WTP (Schkade and Payne 1994). A more quantitative approach to verbal protocols may be an evaluation of segments similar to that of marketing research (Green and Krieger 1991). One could hypothesise that those stating equal WTP for “different goods” in a sequence may differ from other segments in demographics or psychographics.

What remain is to conclude that (most) CV subjects “get it right” even with a public, complex and predominantly passive-use valued environmental amenities. Overall the respondents still show scope insensitivity when a non-optimal elicitation format is used. This study adds to the part of the CV literature concluding that observed insensitivity to scope is due to flaws in survey design and amenity misspecification, not an inherent weakness of the method threatening its theoretical construct validity.

References

Arrow, K.J., Solow, R., Leamer, E., Portney, P., Radner, R. & Schuman, H. 1993. "Report of the NOAA panel on contingent valuation." *Federal Register*, 58: 4601-4614.

Bateman, I.J. & Willis, K.G. 1999 (eds.). *Valuing environmental preferences: theory and practice of the contingent valuation in the US, EU, and developing countries*. Oxford University Press, Oxford.

Bishop, R.C. & Welsh, M.P. 1992. "Existence values in benefit-cost analysis and damage assessment." *Land Economics*, 68(4): 405-417.

Bjornstad, D.J. & Kahn, J.R. 1996 (eds.). *The contingent valuation of environmental resources: methodological issues and research needs*. Brookfield, Vt.: Edward Elgar, Cheltenham, UK.

Boyle, K.J., Bishop, R., Hellerstein, D., Welsh, M.P., Ahearn, M.C., Laughland, A., Charbonneau, J. & O'Conner, R. 1998. "Test of scope in contingent-valuation studies: are the numbers for the birds." Paper presented at the World Congress of Resource and Environmental Economists (AERE/EAERE), June 1998, Venice, Italy.

Boyle, K.J., Desvousges, W.H, Reed Johnson, F., Dunford, R.W. & Hudson, S.P. 1994. "An investigation of part-whole biases in contingent-valuation studies." *Journal of Environmental Economics and Management*, 27: 64-83.

Brown, G.M.Jr. & Shogren, J.F. 1998. "Economics of the Endangered Species Act." *Journal of Economic Perspectives*, 12(3): 3-20.

Carson, R.T. 1997. "Contingent valuation surveys and tests of insensitivity to scope." In Kopp, R.J., Pommerhene, W. & Schwartz, N. (eds.) *Determining the value of non-marketed goods: economic, psychological and policy relevant aspects of contingent valuation methods*. Kluwer, Boston.

Carson, R.T., Flores, N.E., & Hanemann, W.M. 1998. "Sequencing and valuing public goods." *Journal of Environmental Economics and Management*, 36: 314-323.

Carson, R.T., Flores, N.E., Martin, K.M. & Wright, J.L. 1996. "Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods." *Land Economics*, 72(1): 80-99.

Carson, R.T., Flores, N.E. & Meade, N.F. 2001. "Contingent valuation: controversies and evidence." *Environmental and Resource Economics*, 19: 173-210.

Carson, R.T., Groves, T., & Machina, M.J. 2000. "Incentive and informational properties of preference questions." Draft, February 2000. Department of Economics. University of California. San Diego. 47 p.

Carson, R.T. & Mitchell, R.C. 1993. "The issue of scope in contingent valuation studies." *American Journal of Agricultural Economics*, 75: 1263-1267.

Carson, R.T. & Mitchell, R.C. 1995. "Sequencing and nesting in contingent valuation surveys." *Journal of Environmental Economics and Management*, 28(2): 155-173.

Carson, R.T., Wilks, L. & Imber, D. 1994. "Valuing the preservation of Australia's Kakadu Conservation Zone." *Oxford Economic Papers*, 46: 727-749.

Cummings, R.G., Brookshire, D.S. & Schulze, W.D. 1986. *Valuing Public Goods: The Contingent Valuation Method*. Rowman & Allanheld. Totowa, NJ.

Cummings, R.G. & Harrison, G.W. 1994. "Was the Ohio court well informed in its assessment of the accuracy of the contingent valuation method?" *Natural Resource Journal*, 34: 1-36.

Cummings, R.G. & Harrison, G.W. 1995. "The measurement and decomposition of nonuse values: a critical review." *Environmental and Resource Economics*, 5: 225-247.

Cummings, R.G., Harrison, G.W. & Rutström, E.E. 1995. "Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible?" *American Economic Review*, 85: 260-266.

Diamond, P.A., Hausman, J.A., Leonard, G.K. & Denning, M.A. 1993. "Does contingent valuation measure preferences? Experimental evidence." Chapter II in Hausman, J.A. (ed.) *Contingent Valuation: A Critical Assessment*. North-Holland / Elsevier Science Publishers, Amsterdam.

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

Fischhoff, B. & Furby, L. 1988. "Measuring values: a conceptual framework for interpreting transactions with special reference to contingent valuation of visibility." *Journal of Risk and Uncertainty*, 1: 147-184.

Giraud, K., Loomis, J. & Johnson, R. 1999. "Internal and external scope in willingness to pay estimates for threatened and endangered wildlife." *Journal of Environmental Management*, 56(3): 221-229.

Green, P.E. & Krieger, A.M. 1991. "Segmenting markets with conjoint analysis." *Journal of Marketing*, 55(4): 20-31.

Green, C. & Tunstall, S. 1999. "A psychological perspective." In Bateman, I.J. & Willis, K.G. (eds.). *Valuing environmental preferences: theory and practice of the contingent valuation in the US, EU, and developing countries*. Oxford University Press. Oxford.

Greene, W.G. 1997. *Econometric Analysis*. 3. Ed. Macmillan, New York.

Hanemann, W.M. 1994. "Valuing the environment through contingent valuation." *Journal of Economic Perspectives*, 8(4), 19-43.

Hausmann, J. 1993 (ed.). *Contingent valuation. A critical assessment*. Contributions to economic analysis 220. North-Holland / Elsevier Science Publishers, Amsterdam.

Hoehn, J.P. & Loomis, J.B. 1993. "Substitution effects in the valuation of multiple environmental programs." *Journal of Environmental Economics and Management*, 25: 56-75.

Hoen, H.F. & Winther, G. 1993. "Multiple use forestry and preservation of coniferous forests in Norway: a study of attitudes and willingness to pay." *Scandinavian Journal of Forest Research*, 8: 266-280.

Hoevenagel, R. 1996. "The validity of the contingent valuation method: perfect and regular embedding." *Environmental and Resource Economics*, 7: 57-78.

Kahneman, D. 1986. "The review panel assessment: comment". In Cummings, R.G., Brookshire, D.S. & Schulze, W.D. (eds.) *Valuing Public Goods: The Contingent Valuation Method*. Rowman & Allanheld. Totowa, NJ.

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

Kahneman, D. & Ritov, I. 1994. "Determinants of willingness to pay for public goods: a study in the headline method." *Journal of Risk and Uncertainty*, 9: 5-38.

Kemp, M.A. & Maxwell, ?.?. 1993. "Exploring a budget context for contingent valuation estimates." Ch. ?? in Hausman, J.A. (ed.) *Contingent Valuation: A Critical Assessment*. North-Holland / Elsevier Science Publishers, Amsterdam.

Kopp, R.J. 1992. "Why existence value should be included in cost-benefit analysis." *Journal of Policy Analysis and Management*, 11(1): ???-???.

Kriström, B. 1997. "Spike models in contingent valuation." *American Journal of Agricultural Economics*, 79(3): 1013-1023.

Loomis, J., Lockwood, M. & DeLacy, T. 1993. "Some empirical evidence on embedding effects in contingent valuation of forest protection." *Journal of Environmental Economics and Management*, 25: 45-55.

McFadden, D. & Leonard, G.K. 1993. "Issues in contingent valuation of environmental goods: methodologies for data collection and analysis." Ch. ?? in Hausman, J.A. (ed.) *Contingent Valuation: A Critical Assessment*. North-Holland / Elsevier Science Publishers, Amsterdam.

Mitchell, R. C. & Carson, R.D. 1989. *Using surveys to value public goods. The Contingent Valuation Method*. Resources for the future, Washington DC.

Randall, A. & Hoehn, J.P. 1993. "Embedding effects in contingent valuation." Staff Paper No. 93-06 (Draft), Department of Agricultural Economics, Michigan State University, MI, US.

Randall, A. & Hoehn, J.P. 1996. "Embedding in market demand systems." *Journal of Environmental Economics and Management*, 30: 369-380.

Ready, R.C., Navrud, S. & Dubourg, W.R. 2001. "How do respondents with uncertain willingness to pay answer contingent valuation questions?" *Land Economics*, 77: 315-326.

Rollins, K. & Lyke, A. 1998. "The case for diminishing marginal existence values." *Journal of Environmental Economics and Management*, 36: 324-344.

Rowe, R.D., Schulze, W.D. & Breffle, W.S. 1996. "A test for payment card bias." *Journal of Environmental Economics and Management*, 31: 178-185.

SAS 1997. SAS/STAT (R) User's Guide - Version 6. 4th edition. SAS Institute, Inc., Cary., North Carolina. 1848 pp.

Schkade, D.A. & Payne, J.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.

Schulze, W., McClelland, G., Waldman, D. & Lazo, J. 1996. "Sources of bias in contingent valuation." In Bjornstad, D.J. & Kahn, J.R. (eds.). *The contingent valuation of environmental resources: methodological issues and research needs*. Brookfield, Vt.: Edward Elgar, Cheltenham, UK.

Sinden, R. 1999. "Public goods and contingent valuation." In Bateman, I.J. & Willis, K.G. 1999 (eds.). *Valuing environmental preferences: theory and practice of the contingent valuation in the US, EU, and developing countries*. Oxford University Press. Oxford.

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

Smith, V.K. & Osborne, L.L. 1996. "Do contingent valuation estimates pass a 'scope' test?" *Journal of Environmental Economics and Management*, 31: 287-301.

Svedsäter, H. 2000. "Contingent valuation of global environmental resources: test of perfect and regular embedding." *Journal of Economic Psychology*, 21: 605-623.

Thaler, R.H. 1990. "Saving, fungibility and mental accounts." *Journal of Economic Perspectives*, 4: 193-205.

Tversky, A. & Kahneman, D. 1986. "Rational choice and the framing of decisions." *Journal of Business*, 59: 251-278.

Varian, H.R. 1992. *Microeconomic analysis*. 3.ed., Norton & Company Inc., New York.

Veisten, K., Hoen, H.F., Navrud, S. & Strand, J. 1993. "Valuing biodiversity in Norwegian forests: a contingent valuation study with multiple bias testing." Memorandum 7, Department of Economics, University of Oslo, Oslo, Norway.

Acknowledgements

This research was financed by the Norwegian Research Council (partly through the programme “Socio-economic Consequences of Multiple-Use Forestry” and partly through the programme “Valuation and Use of Environmental Costs”). We also thank Olvar Bergland and Michael Farmer for valuable comments on an earlier draft. All remaining errors and omissions remain the sole responsibility of the authors.

Appendix. Test results with full (non-cleaned) samples

Extra-table 4. External scope tests (one-sided). [a]

	Pooled oe – oe-pc	
Null hypothesis	*t* test [b]	MW [c]
$\hat{WTP}\left[A_I^1\right] \geq \hat{WTP}\left[c_{II}^1\right]$	1.49^{*}	1.81^{**}
$\hat{WTP}\left[A_I^1\right] \geq \hat{WTP}\left[b_{III}^1\right]$	4.44^{***}	5.20^{***}
$\hat{WTP}\left[\mathbf{P}_{IV}^1\right] \geq \hat{WTP}\left[A_I^1\right]$	2.97^{***}	1.71^{**}

[a] Based on full (non-cleaned) data sets.

[b] The *t* test gives the student T value assuming equal variances.

[c] MW is the Mann-Whitney rank sum test, and the presented MW test statistic is a normal-approximated T value.

$^{*}p < 0.10$

$^{**}p < 0.05$

$^{***}p < 0.01$

Extra-table 5. Internal scope tests (one-sided). [a]

	Pooled oe – oe-pc	
Null hypothesis	*t* test [b]	MW [c]
$\hat{WTP}\left[A_{II}^2\right] \geq \hat{WTP}\left[c_{II}^1\right]$	1.74^{*}	2.23^{**}
$\hat{WTP}\left[A_{III}^3\right] \geq \hat{WTP}\left[b_{III}^1\right]$	3.61^{***}	3.90^{***}
$\hat{WTP}\left[(b+c)_{III}^2\right] \geq \hat{WTP}\left[b_{III}^1\right]$	1.46^{*}	1.12
$\hat{WTP}\left[A_{III}^3\right] \geq \hat{WTP}\left[(b+c)_{III}^2\right]$	2.86^{***}	2.81^{***}
$\hat{WTP}\left[\mathbf{P}_{IV}^1\right] \geq \hat{WTP}\left[A_{IV}^2\right]$	1.98^{**}	2.69^{***}

[a] Based on full (non-cleaned) data sets.

[b] The *t* test gives the student T value assuming equal variances.

[c] MW is the Mann-Whitney rank sum test, and the presented MW test statistic is a normal-approximated T value.

$^{*}p < 0.10$

$^{**}p < 0.05$

$^{***}p < 0.01$

Extra-table 8. Likelihood ratio test (external scope) based on log-likelihood values from Tobit models. [a]

Null hypothesis	LRT [b] Pooled oe – oe-pc
$\hat{WTP}[A_I^1] \geq \hat{WTP}[c_{II}^1]$	11.9
$\hat{WTP}[A_I^1] \geq \hat{WTP}[b_{III}^1]$	29.1***
$\hat{WTP}[\mathbf{P}_{IV}^1] \geq \hat{WTP}[A_I^1]$	17.8*

[a] Based on full (non-cleaned) data sets.

[b] LRT is the likelihood ratio test, and the test value ($\hat{q}$) is compared to a chi-square distribution with *K* degrees of freedom (*d.f.*), where *K* is the sum of coefficients in the individual regressions, including the intercept, minus the number of coefficients in the pooled sub-samples regression, i.e., *K*=10. The critical value of χ^2 with 10 *d.f.* is ~16.0 at the 0.10 level, ~18.3 at the 0.05 level, and ~23.2 at the 0.01 level.

* $p < 0.10$

** $p < 0.05$

*** $p < 0.01$