Monetary Valuation of Environmental Goods: Alternatives to Contingent Valuation

Baarsma, B.E.

Citation for published version (APA):
Chapter 3
The Contingent Valuation Method

"As the monopolist has no competition to fear [...] he does not make his goods as perfect as would otherwise be the case." (Arndt, p. 49, 1984)

The contingent valuation method (CVM) is the most well-known monetary valuation method and very popular among valuation researchers. CVM is big business, involving large amounts of money. Its uses, and the fee for carrying out such a study, continue to increase at a rapid pace (Knetsch, 1994). Also, the social and financial stakes in the use and abuse of the CVM are very high. CVM possesses some kind of monopoly in the world of monetary valuation methods. As the above quote suggests, like all goods produced by monopolies the CVM is not as perfect as it could have been if competition had been present. Indeed, the method is not at all trouble-free. The popularity and the troubles surrounding the CVM have led to a fierce debate between proponents and opponents of the CVM. This chapter investigates this debate regarding the pros and cons. The next chapter (chapter 4) introduces alternatives to the CVM, that may put the monopoly position of the CVM into perspective.

This chapter is organized as follows. Section 3.1 categorises the different monetary valuation methods. Subsequently, the CVM is described in more detail in section 3.2. The popularity and extensive use of the CVM is the subject of section 3.3. Section 3.4 critically reviews the CVM. Finally, section 3.5 gives some concluding remarks.

3.1 A Classification of Valuation Methods

Various valuation methods are available to put an economic value on environmental goods. Also, various classifications of these methods exist. Bateman and Turner (1993) distinguish two basic approaches, viz. that which values a commodity via a
demand curve and that which does not and therefore fails to provide 'true' valuation information and welfare measures. Another classification comes from Mitchell and Carson (1989) and is based on two characteristics, namely direct versus indirect methods and hypothetical versus observed behaviour methods. Here, a classification is presented that is a combination of the classifications given by the Australian Department of the Environment, Sport and Territories (1995) and by Hoevenagel (1994). Hoevenagel categorises groups of valuation methods on the basis of the process by which these methods retrieve environmental preferences of individuals, viz. stated preference methods and revealed preference methods. The Australian Government report distinguishes a market value approach, a surrogate market approach and a simulated market approach. The resulting classification is presented in table 3.1 below.

**Table 3.1: A classification of valuation methods**

<table>
<thead>
<tr>
<th>REVEALED PREFERENCE METHODS</th>
<th>STATED PREFERENCE METHODS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(observable market data for prices and costs)</td>
<td>(no observable market data for prices and costs)</td>
</tr>
<tr>
<td>MARKET VALUE APPROACHES (derive value from price or costs of an environmental resource)</td>
<td>SIMULATED MARKET APPROACHES (derive value from response to questions in a survey which simulates a market)</td>
</tr>
<tr>
<td>- Change in Productivity</td>
<td>- Contingent Ranking</td>
</tr>
<tr>
<td>- Replacement Costs</td>
<td>- Allocation Games</td>
</tr>
<tr>
<td>- Preventive Expenditure</td>
<td>- Contingent Valuation</td>
</tr>
<tr>
<td>- Relocation Costs</td>
<td>- Conjoint Measurement</td>
</tr>
<tr>
<td>SURROGATE MARKET APPROACHES (derive value from price or costs of surrogate goods or services)</td>
<td>- Welfare Evaluation Method</td>
</tr>
<tr>
<td>- Travel Costs Method</td>
<td>- Well-being evaluation method</td>
</tr>
<tr>
<td>- Hedonic Price Method</td>
<td>- Wage Differential</td>
</tr>
<tr>
<td>- Proxy Good</td>
<td>- Preventive Expenditure</td>
</tr>
</tbody>
</table>

The primary distinction in valuation methods is between stated and revealed preference methods. Stated preference methods are based on preference data that are not observable in the market and that have to be drawn from people's stated responses to hypothetical questions in surveys, whereas revealed preference methods are based on preference data that are observable in the market and that can be revealed from observations of real-world choices. In theoretical terms, this difference between methods comes down to using the uncompensated (Marshallian) demand
Revealed preference methods
The revealed preference methods can be further divided into a market value approach and a surrogate market approach. The former determines the value of environmental resources by using the costs of, or the revenues gained from the effects themselves. These approaches value a benefit as an increase in the revenue or as a decrease in the monetary outlay. Similarly, they value the costs as an increase in the monetary outlay or as a reduction in the revenue. An example of a method included in this approach is the change-in-productivity-technique, which evaluates the effects of a change in the availability, the quality or the quantity of an output. Other examples are the replacement costs technique (the costs made by individuals and society to replace an entire asset, part of an asset, or the quality of an asset), the preventive expenditure method (the money spent to defend the environment, Abdalla et al., 1992), and the relocation costs approach (the costs of relocating an activity from a polluted area).

The surrogate market approaches obtain monetary values from the costs or revenues of surrogates, e.g., effects closely related to the environmental resource involved. When using the hedonic price method, one estimates a property price effect due to a difference in pollution levels (Englin and Mendelsohn, 1991), or, when using the travel costs method, one estimates the economic value of a specific recreational site by looking at the costs of the trips made to this site (e.g., transportation costs, entrance fees, and the opportunity costs of time; Clawson, 1959). Other illustrations of surrogate market approaches are the wage differential method (Clark and Kahn, 1988) by which a change in wages is related to the environmental good, and the proxy good method, which uses the (known) value of a close substitute.

Stated preference methods
All methods of the category of the stated preference methods use surveys to uncover people's preferences. Examples are contingent ranking and allocation games. In case the contingent ranking method is used, people are asked to rank certain specified alternatives. Researchers who use allocation games, ask respondents to allocate sums from a fixed budget among a set of environmental goods.

Another method in the stated preference category is the well-known contingent valuation method (CVM). Since this method forms a major part of the thesis, at least of part I, it will be discussed in more detail in the next section.
In this thesis, three other stated preference methods play a central role: conjoint measurement, welfare evaluation and well-being evaluation. These three methods are not at all well-known methods in the standard environmental valuation literature. They will be discussed in more detail in the next chapter and are only briefly mentioned here.

Conjoint measurement entails the ranking of several vignettes by the respondent. Sometimes, the respondent is also asked to mark the vignettes and to indicate which of the vignettes are acceptable to him or her. The second and third method, welfare evaluation and well-being evaluation, also use stated preferences of the respondents to deduce a valuation for an environmental good. Although these methods are not really market approaches in the sense that goods are offered to be valued, yet they are classified under simulated market approaches, since the goods 'income position' and 'well-being' are valued. In the case of the welfare evaluation method, the respondents are asked to evaluate their income, and a price can be derived by linking this evaluation to environmental variables relevant to the respondent (Van Praag, 1988). The Cantril (1965) question asks respondents to evaluate their personal position on a ladder of life with a scale from 0 to 10. This position represents the respondent's well-being, not his or her welfare. Subsequently, the position on the ladder is related to environmental variables relevant to the respondent.

### 3.2 Introduction to the Contingent Valuation Method

The practical application of the CVM involves six stages (Hanley and Spash, 1993; Bateman and Turner, 1993), namely:

1. preparation
2. survey
3. calculation
4. estimation
5. aggregation
6. appraisal

All of these stages will be discussed briefly below.

**Re 1] Preparation**

First of all the hypothetical market has to be constructed. This means setting up a questionnaire asking basically two kinds of questions: "how much are you willing to pay (WTP) for a welfare gain?" or "how much are you willing to accept (WTA) as
compensation for a welfare loss?”. These questions can be asked in various ways (Jordon and Elnagheeb, 1994), only four of which are mentioned here. By posing open-ended questions (how much are you willing to pay?) a continuous bid variable is produced, whereas by posing take-it-or-leave-it questions (a dichotomous choice format: are you willing to pay $x\$?, where $x$ differs for each respondent) a discrete bid variable is produced. A third elicitation method is the payment card method, which offers respondents a range of WTP values from which they can choose their value. Finally, in an iterative bidding game the interviewer presents an initial bid to the respondent which he or she can reject (the interviewer decreases the initial bid till the maximum WTP is reached) or which he or she can accept (and the bidding game will continue upwards till the maximum WTP is elicited).

Furthermore, the construction of a hypothetical market entails the need to provide information about the quantity and quality change of the good, about who will pay for the good and about who will use the good. Also, the payment vehicle has to be defined, like e.g., taxes, donations or entrance fees.

**Re 2] Survey**
During the second stage, the survey is conducted among a group of respondents who are representative for the population involved.

**Re 3] Calculation**
From the responses, the mean and the median WTP (or WTA) are calculated. In order to calculate a correct value, the so-called protest bids and outliers will have to be omitted. Outliers refer to WTP values that represent an implausibly large part of a respondent’s income, for example more than 10%. Protest bids are typically zero amounts, because these respondents think that others (the polluters) should pay.

**Re 4] Estimation**
In this stage a bid curve is estimated, to investigate the determinants of the stated WTP. Investigating the determinants of the WTA and the WTP bids is useful for assessing the validity of the CVM study. Bid curves open up the possibility of predicting WTP amounts for changes in the level of some environmental variable, provided that the relationships of the variables in the bid curves are stable and significant. Examples of these variables are net monthly income, education level, family size, age, environmental preference and sex.
Re 5] Aggregation
In order to move up from the mean or the median WTP to the total population value, the individual WTP values have to be aggregated over the relevant population. Because a substantial non-response can completely invalidate a carefully designed survey, the WTP response must be reweighted to overcome the problem of misrepresentation of the population.

Re 6] Appraisal
Since the basic aim of the CVM is to elicit true values, it is very important to assess to what degree the result is valid and reliable (cf. table 3.2 below).

<table>
<thead>
<tr>
<th>Table 3.2: Reliability, validity and biases of WTP estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source: adapted from Bateman and Turner, 1993, pp.146-147.</td>
</tr>
</tbody>
</table>

\[
Y = \alpha + \beta X + \varepsilon
\]

where:
\[
Y = \text{measured value of WTP} \\
X = \text{true value of WTP} \\
\alpha, \beta = \text{constants} \\
\varepsilon = \text{residual error}
\]

- **Reliability**: measured by \( \varepsilon \), if \( \varepsilon = 0 \) the estimate is said to be reliable
- **Validity**: measured by \( \alpha \) and \( \beta \), if \( \alpha = 0 \) and \( \beta = 1 \) the estimate is said to be valid
- **Biases**: measured by randomness of \( \varepsilon \). If \( \varepsilon \) is a non-random variable then a bias is likely to be present*

* Biased estimators occur if \( \alpha \neq 0 \) and \( \beta \neq 1 \). Yet, if true values cannot be identified, it is difficult to assess whether the estimate is biased and the extent of the bias. Hence, where the CVM literature speaks of biases, these should be interpreted as potential causes for systematic error.

Reliability refers to the degree to which the variance of WTP responses can be attributed to random error, where holds that the more consistent the results given by repeated measurements, the higher the reliability (Mitchell and Carson, 1989, pp. 211-229). Reliability is, in principle, easy to test; it can be determined by measuring the consistency of the responses over time, or by examining to what extent the variance in the elicited WTP values results from random sources (Loomis, 1989 and 1990; Reilings et al., 1990).

Validity is associated with the degree to which a bid is similar to one that should result if an actual market existed (Mitchell and Carson, 1989, pp. 190-207). Since an actual market does not exist in reality, it is difficult to assess the validity of a particular CVM
study, or any other stated preference study for that matter. Validity must be tested by checking that the resulting bid is similar to one achieved by introducing the kinds of incentives to reveal preference that exist in real markets (Bishop and Heberlein, 1979; Pearce and Turner, 1990). Another possibility is by comparing the resulting bid with the bids from similar techniques, like conjoint measurement (Mackenzie, 1993), or with bids from techniques based on surrogate markets, like the travel costs method (Brookshire et al., 1982; Neill et al., 1994; Smith et al., 1986).

One type of validity that is relatively simple to assess, is theoretical validity (Wierstra et al., 1996). It tests whether the CVM measure conforms to theoretical expectations. For instance, economic theory suggests that the WTP would be larger for large reductions in pollution levels than for smaller ones, and that the WTP increases with the strength of the preference (environmental friendliness) and with the income.

CVM is a stated preference valuation method, and as such it is inherently susceptible to various types of bias. These biases are the common cause of a low validity. In the literature various causes have been identified to account for these biases.

<table>
<thead>
<tr>
<th>Table 3.3: An overview of different sources of bias</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>General biases</th>
<th>Procedural biases</th>
<th>Instrument biases</th>
</tr>
</thead>
<tbody>
<tr>
<td>* strategic bias</td>
<td>* sampling bias</td>
<td>* payment vehicle bias</td>
</tr>
<tr>
<td>* hypothetical bias</td>
<td>* compliance bias</td>
<td>* starting point bias</td>
</tr>
<tr>
<td>* information bias</td>
<td>* interviewer bias</td>
<td>* relational bias</td>
</tr>
</tbody>
</table>

Strategic bias occurs if the respondent understates his or her value for an environmental good on the assumption that others will pay for its provision (free-riding) or if the respondent overstates his or her WTP in an effort to raise the mean and thereby ensure provision.

Hypothetical bias concerns the question whether a respondent’s declared intentions (stated WTP) can be taken as meaningful guides to his or her actual behaviour (true value). Hypothetical bias might occur if the very fact that respondents are asked for valuations in a hypothetical market makes their responses differ systematically from real cash (‘true’) values.
Information bias arises if the quality and quantity of the information about the hypothetical market in the survey affects the responses received (Bergstrom et al., 1989). Hoevenagel (1994) found that people express higher and more accurate WTP bids as they receive more information. Respondents' bids may be influenced by either the hypothetical market or the commodity-specific information given to them in the survey, which indicates that the WTP or the WTA bids are endogenous to the valuation process. As the respondent is often unfamiliar with the hypothetical market, the large amount of information may be too much to absorb. The order in which the information is presented and in which the questions are asked, can also have a considerable effect on the WTP value (for an example in which the question order did affect the WTP bid, see Diamond and Hausman, 1994; and for one in which it did not matter, see Blackburn et al., 1994).

An example of sampling bias is self-selection bias, which implies that respondents with certain characteristics, like strong environmental preferences or a high education level, have a higher response rate than those with weaker preferences and a lower education level, and are thus overrepresented in the sample (for other examples, see Edwards and Anderson, 1987). Furthermore, it can be argued that there are 'good respondents', who state a higher WTP just to please other people (compliance bias), or that the very character of the interviewer or the interview may influence responses (interviewer bias).

Sometimes the payment vehicle affects the WTP. For instance, some Americans may hate taxes so much that they will not be willing to pay anything, regardless of the good in question (payment vehicle bias). Starting point bias occurs when respondents base their WTP value on the starting point in a dichotomous choice survey or in a bidding game and, similarly, range bias occurs when respondents base their WTP values on the range of amounts displayed on the payment card. Finally, relational bias is present when respondents base their WTP value on the prices of perceived related goods.

### 3.3 Use and Popularity of the Contingent Valuation Method

The popularity of contingent valuation is considerable. Several thousands of CVM studies have been conducted. Most textbooks about environmental economics deal with the CVM, journals publish piles of articles about the CVM, and most valuation papers presented at conferences are based on the CVM. To a large extent, this popularity is caused by legislative developments in the United States (US).
US legislative developments

In February 1981 Ronald Reagan, the then president of the US, issued Executive Order No. 12291. This order required that all major new regulations be subjected to a cost-benefit analysis before they could be enacted (Smith, 1984). It was the first order which explicitly required cost-benefit analysis for new regulations. Before the proclamation of the No. 12291 Order, earlier requirements for the inclusion of environmental damages concerned physical measures and environmental impact assessments, similar to current legislation in the European Union (Hanley and Spash, 1993). Because Order No. 12291 has placed a strong emphasis on the measurement of costs and benefits of environmental regulation, the use of valuation methods was stimulated greatly.

In addition to Reagan's Order, legislation such as the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980 has also encouraged the measurement of environmental benefits and damages. The CERCLA established the legal and procedural framework for the Executive Departments to act as trustees for certain protected resources and to prosecute public claims against potentially responsible parties for damages from injury to natural resources as a result of oil or other hazardous substance spills (Bockstael and Strand, 1994). The courts determine the compensation to be paid to the public by the polluter. According to the final rule promulgated by the Department of the Interior (in 1986), contingent valuation has been recognised as an approved method for measuring benefits and damages under the CERCLA.

Finally, the idea of using contingent valuation to estimate harm to natural resources has been stimulated by a decision of the US Court of Appeals for the D.C. Circuit in the State of Ohio versus the US Department of the Interior, in 1989. The decision encourages the measurement of the full damages to natural resources, including non-use components, and therefore the use of the CVM, as it appears to be the only method capable of incorporating non-use values (Arrow et al., 1993; Cummings and Harrison, 1994).

This advantage of the CVM over other techniques, has strengthened its monopoly position even more. However, the argument that the CVM is the only method that can measure non-use values and, consequently, has something like a natural monopoly, is not convincing, because other valuation methods are also capable of measuring non-use values, as long as they are stated preference methods (Baarsma, 1997a, pp. 279-280). In order to elucidate this point, the concepts of non-use value, use value, option value and bequest value are very briefly discussed here.
Measurement of non-use values

Non-use value, or existence value, is unassociated with actual use by an individual. It is motivated by some form of altruism and, therefore, it is still an anthropocentric concept (Mazotto and Kline, 1995). An example is the existence of humpback whales; even if we do not 'use' them (e.g., we do not eat, film, or see them), we could still value their existence. Non-use value can be distinguished from regular use value, from option value, and from bequest value (Pearce and Turner, 1990). Use values derive from the actual use of the environment. This actual use can be direct (timber revenues from a woodland) or indirect (fixing carbon dioxide levels when more trees are planted). Option values relate to the potential use of an environmental good at a later date by the person who expresses the value, or by someone currently known to the valuer. It is a kind of insurance premium to ensure the future availability of the environmental asset. Bequest values relate to the potential use by an individual's descendants, or, more generally, by future generations.

Since revealed preference methods, like the travel costs method and the hedonic price method, are based on observed behaviour in actual markets where goods are traded that individuals have a user-relation with, these methods are not capable of measuring non-use values. Non-use values simply are not reflected in market transactions, because non-users do not buy (or sell) 'goods with non-use value' in an actual market (Freeman, 1993). However, some revealed preference methods, like the proxy good method, can account for some non-use values. For instance, a donation to an environmental organization like Greenpeace, in aid of their campaign against the hunting of seals in Norway, may be considered partly as the non-use value for Norwegian seals. Partly, since, on the one hand, a portion of each donation is used for user purposes (magazines, field trips and administration) and since, on the other hand, aggregate donations are likely to be less than the total non-use values, because of the free-rider behaviour of (potential) contributors.

Unlike revealed preference methods, stated preference methods use hypothetical markets where all kinds of goods can be surveyed, and thus also goods that individuals do not have a user-relation with. In fact, all stated preference methods are capable of measuring the total value of a good, including non-use values (Baarsma, 1997a). So, in this respect the stated preference methods described in the next chapter, viz. the conjoint measurement, the welfare evaluation and the well-being evaluation method, are full alternatives to the CVM. Nevertheless, the CVM has one advantage over these other stated preference methods. As the CVM is a direct method, it is easier to divide the total value into a use component and a non-use component. The alternatives are all indirect methods, in the sense that the valuation is
deduced from an ordering of vignettes or from an evaluation of income or well-being, which complicates the division into two separate value components. When using indirect valuation methods, it is impossible to ask respondents how much of their WTP is related to use values and how much is related to non-use values.

However, the validity of such a division into value components is questionable. Cummings and Harrison (1995) contend that there is no operationally meaningful way in which one might decompose total value into use value and non-use value. The sole possibility is by estimating values for groups of users and relate these to use values, and by estimating values for groups of non-users and relate these to non-use values. However, the valuation of some non-users might be too high if they anticipate on the future use of the good in question (in other words, if they include an option value in their valuation). Moreover, Diamond and Hausman (1993) argue that it is not necessary to divide total value into components, since it is total value only that matters in cost-benefit analyses.

Notwithstanding the fact that other valuation methods belonging to the stated preference group can measure total value, the CVM does not have to fear their competition. The CVM even is so well-established and accepted in the US that the results of one particular CVM study, namely the Exxon Valdez study, have led to new legislation.

**CVM and the Exxon Valdez oil spill**

On March 24, 1989, the tanker Exxon Valdez, carrying more than 50 million gallons of crude oil, ran aground and ruptured its tanks on Bligh Reef in the Prince William Sound in Alaska. The oil spill that followed was the largest tanker oil spill in the history of the US. The oil killed thousands of wild animals and has affected many others, and also polluted a lot of the marine plants, micro-organism, coasts and water.

The state of Alaska filed a suit against the Exxon Corporation. The state claimed compensation for a wide range of natural resource injuries. A jury decided that the tanker's owner Exxon and its captain had been reckless. So, they had to pay punitive damages of up to $5 billion (NRC, 17/7/94). The state of Alaska and the federal government also brought a $3 billion claim for the loss of natural resources. Less than 10% of the claim represented the market value of lost animals and lost time, the rest, i.e. $2.8 billion, represented non-use value and was based on CVM estimates.⁴
Exxon did not actually pay the $3 billion. Exxon felt they did not have to pay this amount because of the fact that, under the laws at the time of the oil spill (1989), only direct, use-related, damages could be compensated. The state of Alaska and the federal government did not want to gamble that the $3 billion CVM-based claim would hold up in court, and therefore they settled with Exxon for $1 billion.

It was this Exxon Valdez story, with the CVM-based claim of $2.8 billion of non-use values, that provided the impetus to administer a new law known as the Oil Pollution Act in 1990. Under this law it is possible to ask for compensation payments, not only with regard to the direct damage (related to use values) but also with regard to the indirect damage (related to non-use values) caused by oil spills.

**Popularity of the CVM in the rest of the world**

By now the CVM is a well-respected method among regulators, (economic) scientists and jurists, at least in the US. Although the CVM is not as popular in the rest of the world, most countries acknowledge the importance of monetary valuation of the environment. For instance, 170 countries have agreed to the Rio Declaration on Environment and Development, which recommends the establishment of a process for focusing on the pricing and valuation of environmental goods and services (UNCED, 1992, 8.37 (a) to (c)).

In Australia the popularity of the CVM has declined since 1992. The governmental Resource Assessment Commission conducted a CVM study to measure the costs and benefits of a large-scale project to develop a nature reserve, which implied, among other things, the extraction of tin (Carson et al., 1994). The results very strongly rejected the extraction alternative: extrapolated to the whole Australian population, the total willingness to pay to conserve the area against mining ranges from $435 million to $1,170 million (Australian dollars), which is greatly in excess of the net benefits from mining ($102 million). This outcome stimulated a severe discussion about the usefulness of the CVM. This discussion was fuelled, among other things, by the fact that respondents who lived further away from the nature reserve were willing to pay more than respondents who lived closer. After having consulted some American CVM experts, the Australian government has decided that it will not officially use the CVM in the case of large-scale environmental projects (Pearce, 1993, pp. 73-74). For small-scale projects the CVM is still considered useful (Beder, 1993, pp. 50-51).

Meanwhile, in Europe, development of the CVM has been relatively slow. Article 130 R of the legislation of the European Union (EU) includes a weak version of the cost-benefit analysis requirement contained in the US Order No. 12991. It states that the
EU will take into account the benefits and the costs which arise from her actions or lack of action (Hoevenagel and Opschoor, 1990). More recently this requirement has been reinforced, since the European Union now suggests to analyse the costs and benefits of proposed new rules (European Trends, 1994). Within Europe the CVM is practised mainly in the Scandinavian countries and Great Britain (UK). For instance, the UK government requires the valuation of benefits of projects to be undertaken, and benefits to be compared with costs, as a prerequisite to sanctioning public investment and expenditure (Willis, 1994).

In the Netherlands, the method has been used by some economists but not by governmental agencies. The Dutch government’s interest in monetary valuation methods is not so overwhelming as in the US (Nieuwsbrief milieu en economie, 1999). The reason is that most of the new environmental regulations are included in the National Environmental Policy Plans, and because these plans are accepted by majority decisions, most officials do not deem it necessary to show that the benefits of each measure outweigh the costs (RMNO, 1994, p. 30). On the other hand, one of these Environmental Policy Plans (the NMP II) states on several occasions that the costs of environmental degradation should be internalised in product prices, and thus recommends the use of monetary valuation (ibidem, p. 27).

In conclusion, monetary valuation and particularly the CVM appear to be well-known. Apart from the scientific interest and practice, governments get to know the method, meaning that they have already used it, are contemplating its use or are showing an interest in it.

### 3.4 A Critique of the Contingent Valuation Method

From the previous section it is obvious that the CVM is very popular. However, the method has a lot of critics. Many of these critics were present at a symposium on environmental valuation held by the petroleum industry in 1992 (cf. Hausman, 1993). This symposium coincided with the public comment period for the rule-writing process of the Oil Pollution Act (Bockstael and Strand, 1994). By then, the $2.8 billion result of the CVM study conducted by Carson et al. (1992) to determine the losses related to non-use values in the Exxon Valdez oil spill, was already known. The petroleum industry dreaded the incorporation of non-use values in damage assessments, as that could substantially increase liability claims. The economists and other scientists present at the Symposium warned against the use of the CVM in
natural resource damage cases, and argued that only use values or market values should be incorporated.

The discussions from the petroleum industry symposium prompted the appointment of the National Oceanic and Atmospheric Administration (NOAA) panel, headed by two Nobel Laureates (Arrow and Solow), in order to provide an unbiased assessment of the validity of CVM measures of non-use values. The NOAA panel provided an extensive set of guidelines for CVM survey construction, administration and analysis. In the panel’s view, "the more closely the guidelines are followed, the more reliable the results will be" (Arrow et al., 1993, p. 4609). Since the panel found that the CVM, if appropriately conducted, could convey useful information, the debate has become fiercer and more charged. Perhaps that is logical, given the fact that the social and financial stakes in the use and abuse of the CVM are very high, even more so since the Oil Pollution Act was passed.

In this section, some of the most frequently disputed issues surrounding the CVM are discussed, namely embedding effects, endowment effects, overestimation and aggregation. Since most of these issues can be traced back to the discussion of neoclassical assumptions about individual behaviour in chapter 2 (section 2.1.2), the various critiques in the next sections may be overlapping parts of chapter 2.

3.4.1 Embedding Effects

The embedding phenomenon encompasses situations where different, but similar, samples of respondents are asked about their WTP for public goods that are identical except for their scale. For instance, the 'inclusive' (all encompassing) good, e.g., saving nature, incorporates several 'embedded' goods, e.g., saving the animals and saving the seal. Kahneman and Knetsch (in: Kahneman, 1986) were the first to demonstrate such an effect.

Kahneman and Knetsch's results

Kahneman and Knetsch found that the expressed willingness of Toronto residents to pay increased taxes to prevent the drop in fish populations in all Ontario lakes (an inclusive good) was not significantly higher than their WTP to preserve fish stocks in only a small area of the province (an embedded good).

Later, in 1992, they conducted a more extensive survey in which they valued a good ABC (environmental services), an embedded good AB (improve disaster preparedness) and an even more embedded good A (improve rescue equipment and
personnel). They asked three groups to participate. Group 1 was asked to value ABC and then to allocate values to AB and A, group 2 was asked to value AB and then to allocate a value to A, and group 3 was merely asked to value A. They found significant evidence that the CVM suffers from an embedding effect. The results are presented in table 3.4.

**Table 3.4: Mean WTP (in US$) for an inclusive good and two embedded goods**

Source: Kahneman and Knetsch (1992a), p. 61

<table>
<thead>
<tr>
<th>Good</th>
<th>Group 1 (n=66)</th>
<th>Group 2 (n=78)</th>
<th>Group 3 (n=74)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental services (ABC)</td>
<td>$135.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Improve disaster preparedness (AB)</td>
<td>$29.06</td>
<td>$151.60</td>
<td>-</td>
</tr>
<tr>
<td>Improve rescue equipment and personnel (A)</td>
<td>$14.42*</td>
<td>$74.65</td>
<td>$122.64</td>
</tr>
</tbody>
</table>

* K&K incorrectly reported this value as $14.12. Nickerson (1993) gave the correct value.

Kahneman and Knetsch tested for three types of embedding effects, as is shown in table 3.5 below. The first is referred to as perfect embedding, which occurs if the same WTP is observed for an embedded commodity and an inclusive commodity (viz. goods A, AB and ABC). According to their analysis, perfect embedding occurred, that is, the WTP for the public good ($135.91, $151.60 and $122.64) is not significantly affected by the inclusiveness of the good.

**Table 3.5: Three types of embedding in the Kahneman and Knetsch study**

<table>
<thead>
<tr>
<th>Type</th>
<th>WTP(ABC/group 1) = WTP(AB/group 2) = WTP(A/group 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect embedding</td>
<td>WTP(AB/group 1) &gt; WTP(AB/group 2) or WTP(A/group 1) &gt; WTP(A/group 2) &gt; WTP(A/group 3)</td>
</tr>
<tr>
<td>Regular embedding</td>
<td>WTP(A/one-time payment) = WTP(A/every year for a period of 5 years)</td>
</tr>
</tbody>
</table>

The second type of embedding, called regular embedding, arises when the WTP assigned to a good is lower when derived from the WTP value of an inclusive good than when valued on its own. In this study, there is a significant effect of the position in the embedding structure on stated WTP ($29.06 versus $151.60 for good AB and ranging from $14.12 to $122.64 for good A).
The last type of embedding is temporal embedding, which exists if respondents cannot discriminate between a one-time payment and a long-term commitment (a series of payments) for a good. The hypothesis of temporal embedding was not rejected.\(^6\)

Smith (1992) and Harrison (1992) raised several doubts about this Kahneman and Knetsch study of embedding.\(^7\) However, many other CVM studies also report embedding effects. Perhaps the most notable of these studies is a series of tests carried out by a group of prominent economists supported by the Exxon Corporation (reported in: Hausman, 1993), like the study of Desvogues et al. (1993) discussed earlier (section 2.1.2.d in chapter 2), and the study of Diamond et al. (1993). They studied, among other things, whether respondents are willing to pay significantly more to preserve one, two or three wilderness areas (perfect embedding). The sample mean for three areas is $45, whereas the sample means for the three areas individually are $50, $30 and $37. On the basis of these figures, they concluded that perfect embedding is present.

There are also CVM studies that do not suffer from an embedding effect. For instance, Bateman et al. (1994) studied the WTP for landscape changes in the Yorkshire Dales (UK), arising from a change in agricultural practice, and that for landscape and habitat changes in the Norfolk Broads (UK), arising from increased flood risk. These researchers found no evidence of the embedding effect; in their study people were able to see the difference between several levels of landscape change (from a mild change to a radical change). Another example is a study by Carson et al. (1996), who tested and accepted the hypothesis that the WTP for a recovery plan to mitigate reproductive problems of two species within 5 instead of 15 years is significantly lower, than the WTP for a plan to accomplish this objective for four species within 5 instead of 50 years.

Nevertheless, the fact that some studies do not suffer from an embedding effect does not imply that CVM results are, in general, insensitive to embedding. Actually, the number of studies with embedding results is so large that embedding effects can no longer be overlooked (Brown et al., 1995b; Diamond et al., 1993).

**Explanations of embedding effects**

Embedding effects point to the following three problems:

1. mental account bias
2. part-whole bias
3. warm glow
First of all, respondents may have difficulty taking into account their available income and other demands when making their WTP bids. Hanley and Spash (1993) speak of mental account bias, which means that respondents have some account for environmental issues (the environmental budget $B_i$), but instead of stating their true WTP for an environmental good $x$ ($B_{ix}$), they bid an amount $B_{ix}$, so that $B_{ix} > B_{ix} > B_{ix}$. In order to make the results more consistent with the theory, it may be necessary to explicitly introduce mental accounts into CVM surveys, by asking some initial questions about the respondent's total yearly budget for all environmental issues, including those donations and subscriptions that he or she might already have made.

The second problem related to embedding effects, is the fact that respondents may have difficulty in separating one aspect from a larger asset. This is referred to as part-whole bias, where respondents value a larger entity than the researcher's intended good.

A third problem related to embedding effects is the fact that people may give charitable contributions largely for the pleasure of giving, instead of expressing a preference for an environmental good. Earlier (chapter 2, section 2.1.2.f), this tendency was referred to as warm glow or moral satisfaction. In short, the warm glow or moral satisfaction associated with contributions to an inclusive or embedded good extends with little loss to any subset of that good.

If the respondents' behaviour is indeed subject to mental account bias, part whole bias and the warm glow theory, the resulting WTP will be an overestimation of their true value (cf. section 3.4.3 below). Several experiments show these overestimations. For instance, the WTP for a public good can vary by a factor of one hundred or more, depending on whether that good is considered on its own or as part of a much larger bundle of goods (regular embedding; Kemp and Maxwell, 1993).

According to the 'believers' in the CVM, embedding is caused by improperly designed surveys (Carson et al., 1996). To them, the prevalence of the embedding effect depends to a large extent on whether the resource valued is familiar or unfamiliar to the respondent. Furthermore, they claim that embedding effects as experienced in CVM studies are not necessarily inconsistent with the theory of economic choice.

Economic theory assumes that the availability (or absence) of substitute goods matters for the valuation of a good. When sequentially valuing environmental goods that are close substitutes, the second and the following goods will be valued lower than they would have been if valued independently. This is caused by the fact that the former
goods act as close substitutes for the latter goods. This means that the presence of (close) substitutes for the environmental goods and the sequence in which the goods are valued in a CVM survey, influence the stated WTP values. Hence, in some instances it can be argued that embedding simply indicates diminishing marginal utility of the good being valued.

For example, Hoehn's (1991) empirical results indicate such significant substitution effects in valuing environmental conditions across different geographically protected regions. He claims that these substitution effects are not merely due to diminishing marginal valuation in consumption, but also to the fact that, as the total size of the area to be protected increases, the likelihood of substitution effects between sub-areas within the larger area increases too. In addition, he states that, as the absolute dollar value to protect larger areas increases, substitution effects with other goods become more intense (see also Loomis et al., 1993). According to Hoehn, part of the embedding problem (he refers to it as an overestimation bias) can be solved by correcting for these substitution effects.

However, a large part of the embedding problems is not consistent with economic theory. After all, rational respondents are supposed to recognise the implicit consequence of stating a WTP in a valuation question, or, in other words, respondents should know they could be asked to value other environmental goods, and they should keep those other potential valuations in mind when valuing the good concerned. Psychological research shows that respondents do not obey this economic assumption; they act according to the principle of "what is out of sight, is out of mind" (Knetsch, 1994). A solution could be to present in the survey as much information as possible about close substitutes and (environmental) budget constraints. However, this solution could never solve the problem of embedding, since it is virtually impossible to name all relevant substitutes. Moreover, the risk of an overload of information prevails. Notwithstanding these facts, the NOAA panel on contingent valuation recommends that respondents should be "forcefully" reminded about substitutes and their budget constraint prior to posing the WTP question. Most CVM researchers agree with this recommendation (Hoehn, 1991; Hoehn and Randall, 1989).

But the evidence on this point is not unanimous. A study by Loomis et al. (1994) about the benefits of reducing fire hazards to old-growth forests in Oregon (US) suggests that reminders of substitutes do not significantly influence CVM estimates. The same applies to a study conducted by some of the Exxon economists (Diamond et al., 1993), which shows no significant differences in mean WTP, regardless of the number of substitute goods mentioned and valued in the survey. On the other hand, a study
conducted by Cummings et al. (1994) suggests that reminders of substitutes do influence valuation. Somewhere in between lie the results of a study by Neill (1995), who compared the WTP of three questionnaires A, B and C. Questionnaire A only provides information on a public good without discussing possible substitutes, B describes 8 public goods and then asks for the WTP for one of them and, finally, questionnaire C asks for the WTP for all 8 goods. There is no significant difference between the mean WTP in questionnaires A and B, but the mean WTP of B and C differs significantly indeed. Neill concludes that simply reminding respondents of possible substitutes is not enough.

To sum up the literature, the perception of substitutes may play some role in valuations, but there appears to be little empirical support for believing that such perceptions are an exhaustive and plausible explanation for reported cases of embedding.

The other part of the recommendation of the NOAA panel entails that, when answering CVM questions, respondents should be reminded of their budget constraint. This recommendation was prompted by the idea of mental account bias. According to this idea, embedding effects would be found only in WTP judgements and not in WTA judgements. After all, the WTA is not limited by an account of spendable income, whereas the WTP is. Nonetheless, Baron (1996) found just as much embedding effect in the WTA as in the WTP, which speaks against the budget constraint hypothesis.

### 3.4.2 Endowment Effects

A second problem related to the CVM is the fact that responses to CVM surveys show a persistent difference between the willingness to pay and the willingness to accept compensation measures for economic losses (Gregory, 1986). This WTP/WTA disparity is generally referred to as the endowment effect. Endowment effects occur if respondents overvalue an asset due to the possession of it or, in other words, due to loss aversion. The precise magnitudes of the WTA/WTP difference vary across studies, but ratios of four to fifteen times the actual value or more are not uncommon (Milgrom, 1993). An early example of the WTA/WTP disparity is given in a paper by Brookshire et al. (1980) about the annual value of the right to hunt elk, as influenced by the hunting environment and the frequency of encounters with elk. They use an iterative bidding technique to obtain estimates of the WTP for increments in the quality of wildlife-related amenities and the WTA to permit decrements. Depending on the range of the frequency of encounters per hunting day (0.1-1; 1-5; 5-10), the results exhibited endowment effects; the WTP and the WTA for the first range were $43.64
and $68.52, respectively, for the second range $54.06 and $142.60, and for the third range $32.00 and $207.07.

At first glance, it seems logical that the WTA is much higher than the WTP. The use of the WTA measure could lead to enormous income effects, since there is no upper limit on what a person could ask for as compensation for foregoing the improvement or accepting the loss, whereas the WTP is constrained by the individual’s income. On the other hand, the WTA does depend on income, despite the absence of a budget constraint. For instance, a rich person will ask for a higher monetary compensation than a poor person, because his or her marginal utility of income is lower.

Apart from this interpretation, several other explanations have been suggested to resolve the problem of endowment effects. These explanations can be divided into an economic and a psychological category.

**Economic explanations of the endowment effect**

Three economic explanations for endowment effects will be reviewed:

1. a traditional neoclassical explanation
2. an explanation based on a more realistic market setting
3. an explanation based on the distinction between out-of-pocket costs and opportunity costs

Firstly, according to Hanemann (1991) neoclassical theory can explain endowment effects by considering income effects as well as substitution effects. The work of Hanemann is based on preceding work by Willig (1976) and Randall and Stoll (1980). Willig’s calculations show that, in the case of price changes, the difference between the compensating variation and the equivalent variation will in most cases be less than 10%, depending on the income elasticity of the demand. Since most environmental changes concern quantity changes rather than price changes, Randall and Stoll show that, for quantity changes and small income effects, the Willig result is still valid. This is what we saw in chapter 2 (figure 2.2 in section 2.1.1), specifically that economic theory indicates that the willingness to accept a welfare loss (the compensating variation COM: area $z_1$KLZ$_2$ in the lower panel) will be somewhat greater than the WTP to avoid this loss (the equivalent variation EV: area $z_1$MNZ$_2$) and, similarly, for a welfare gain, that the WTA (now measured by EV: area $z_1$KLZ$_2$) is greater than the WTP (measured by COM: area $z_1$MNZ$_2$). Summarising, the WTP is less than the WTA for either gains or losses, but the derivation of these measures changes (i.e., COM or EV).
When confronted with the growing evidence of endowment effects that could not be explained with the analyses by Willig and by Randall and Stoll, Hanemann (1991) showed that welfare theory indicates that one should expect large differences between the WTP and the WTA measures, because in general their required conditions are seldom met. He argues that:

“for quantity changes, there is no presumption that WTP and WTA must be close in value and, unlike price changes, the difference between WTA and WTP depends not only on an income effect but also on a substitution effect.” (p. 635)

Hanemann stated that if there is a low elasticity of substitution between the good being valued and market goods, i.e., if the good being valued is considered unique in some respect by all respondents, and if there are no income effects, the WTA and the WTP should be equal.9

Still, a study conducted by Adamowicz et al. (1993) shows that the hypothesis that the presence of substitutes decreases the difference between the WTA and WTP measures, must be rejected. Also, simulated experiments involving commonplace goods have yielded substantial differences between the two measures (Kahneman et al., 1990; Knetsch and Sinden, 1984 and 1987; Knetsch, 1989), which cannot be explained by Hanemann's analysis.

A second economic explanation for endowment effects is given by researchers who suggest that the disparity between the WTA and the WTP disappears when proper incentives, that is, market-like incentives, are established for people to state their true values (Blomqvist, 1989; Tietz, 1992). Moreover, markets contain built-in mechanisms whereby buyers (WTP) and sellers (WTA) learn about the commodity they are trading, whereas the hypothetical once-only character of a CVM question does not entail such a learning mechanism. For instance, Knez et al. (1985) show that if the trading of a good is repeated over several consecutive market periods, the difference between the WTA and the WTP decreases.

In another study, Brookshire and Coursey (1987) compare value elicitation for an increase (WTP) in the base plan from 200 to 225 and from 200 to 250 trees, and for a decrease (WTA) from 200 to 175 and from 200 to 150 trees. Two elicitation procedures are used for this purpose, namely a CVM survey and a more market-like Smith auction laboratory process (SAL).10 The results, shown in table 3.6, indicate that the CVM setting indeed yields an enormous endowment effect: the WTA is approximately 75
times higher than the WTP. On the other hand, the SAL setting produces much smaller WTP/WTA disparities. Moreover, the disparities slightly decrease after five bid series, which could point to a learning effect (the WTA/WTP ratio is 3.9 to 8.1 times for the first bids, and 2.4 to 7.4 times for the final bids).

| Source | adapted from Brookshire and Coursey, 1987, p. 561 |

<table>
<thead>
<tr>
<th>CVM</th>
<th>SAL</th>
<th>Initial bids</th>
<th>Final bids</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTA</td>
<td><em>25 trees</em>¹</td>
<td>$855.50</td>
<td>$28.63</td>
</tr>
<tr>
<td></td>
<td><em>50 trees</em>²</td>
<td>$1,734.40</td>
<td>$67.27</td>
</tr>
<tr>
<td>WTP</td>
<td><em>25 trees</em>¹</td>
<td>$14.00</td>
<td>$7.31</td>
</tr>
<tr>
<td></td>
<td><em>50 trees</em>²</td>
<td>$19.40</td>
<td>$8.33</td>
</tr>
</tbody>
</table>

¹ WTA/WTP for a decrease(WTA)/increase (WTP) in the base plan from 200 to 175 trees.
² WTA/WTP for a decrease(WTA)/increase (WTP) in the base plan from 200 to 150 trees.

However, Shogren et al. (1994) show that for non-market goods with imperfect substitutes (such as saving a bird species from extinction, or the prevention of oil spills in certain areas), the divergence between the WTA and the WTP values is persistent, even with repeated market participation and full information on the nature of the good.

Finally, the third economic explanation for endowment effects is suggested by Thaler (1980). He argues that, in the case of a welfare loss, the WTA measure is based on the loss of a good currently held and can thus be seen as out-of-pocket costs, whereas the WTP measure is based on foregone opportunities and can therefore be treated as opportunity costs. Since opportunity costs have less weight in people’s decisions than do out-of-pocket expenses, it is not surprising that the WTA exceeds the WTP. This kind of behaviour is not uncommon in the real world, as the following example from Gregory (1986) illustrates. Many people who paid £10 for a theatre ticket would refuse to sell it for £25 at the night of the performance, but would also refuse to pay £25 for a new one if the ticket had not yet been bought.

**Psychological explanations of the endowment effect**

Psychological explanations for the endowment effect encompass prospect theory and the theory of cognitive dissonance.
Prospect theory was developed by Kahneman and Tversky (1979). Prospect theory does not bridge the gap between the WTA and the WTP, it merely explains it. This theory is a critique of expected utility theory. Expected utility theory cannot explain why people’s preferences reverse when the same choice problem is presented as a positive prospect or as a negative prospect. For example, 80% of a group (n=95) chose $3,000 over a 80% chance to win $4,000, whereas 92% of the same group of people chose a 80% chance to lose $4,000 over a certain loss of $3,000. Prospect theory, on the other hand, can explain this behaviour, by proposing a value function that is (Kahneman and Tversky, 1979, p. 279):

"(i) defined on deviations from a reference point
(ii) generally concave for gains and commonly convex for losses
(iii) steeper for losses than for gains"

This value function encompasses the idea that the disutility of giving up an object is greater than the utility associated with acquiring it. Therefore, it can explain the difference between the WTA and the WTP: people require far more compensation to give up a public good (WTA) than they are willing to pay to acquire it (WTP).

Related to this idea of prospect theory is the notion of framing effects as introduced in chapter 2 (section 2.1.2.f), when the idea was discussed that preferences are not pre-established but are formed during a CVM survey. In the case of endowment effects, this implies that much higher values are elicited if you frame a valuation question as a loss instead of as a gain (Dubourg et al., 1994).

The second psychological explanation for endowment effects is based in the theory of cognitive dissonance developed by Festinger (Antonides, 1991, pp. 195-201). Cognitive dissonance theory states that an inconsistency between two or more cognitive elements in an individual’s mind will motivate the individual to decrease the dissonance and avoid situations and information which would be likely to increase the dissonance. When confronted with an environmental loss, respondents experience a dissonance between the loss of the environmental good on the one hand, and the appreciation of that good and feelings of responsibility for future generations and for the environment in general on the other hand. In order to decrease this dissonance between the different cognitive elements, respondents might refuse to answer any valuation questions (which would explain the relatively large non-response in WTA surveys), or they might state improbably large WTA amounts (Antonides and Knetsch, 1995). Furthermore, it is assumed that, when respondents are confronted with a loss and are asked how much they are willing to pay to avoid the loss (WTP), the resulting
cognitive dissonance is far less than in WTA situations, thereby explaining the large WTA/WTP disparities.

**WTA or WTP?**

So far, the empirical problems surrounding the WTA and the WTP have not been settled (Freeman, 1993). The choice between a WTA question format versus a WTP question format is based on, among other things, the definition and distribution of property rights. The WTP measure implicitly assumes that the agents involved do not have previous rights to the resource being valued. If respondents would have had previous rights (from a philosophical or legal standpoint), the appropriate measure of net economic value would be the WTA measure. Notwithstanding the fact that, in general, the WTA will be the correct measure, in most cases the WTP measure is used. Here, three often expressed reasons for this inclination for the use of the WTP measure are presented.

Firstly, WTP measures are chosen because they seem to correspond more closely to most of the market exchanges people make, and therefore involve people in a more familiar situation. Secondly, the WTP is chosen over the WTA because the former provides a conservative choice, so that the overestimation due to hypothetical bias can be corrected (Arrow et al., 1994; also see the next section). This is, of course, a rather inexact, unscientific argument, since the magnitudes of overstatement and understatement are not known. Still, CVM practitioners defend their preference for conservative estimates, because these lower values tend to be accepted more easily by lawyers, business people and politicians, who are after all the people who make use of and sponsor CVM studies.

Endowment effects have serious policy implications as well as theoretical implications. Some of the policy implication of wrongly using the WTP instead of the WTA are an understatement of losses, setting the standard at inappropriate levels, a bias on policy selections, the encouragement of environmentally degrading activities, and the discouragement of mitigation efforts (Knetsch, 1990).

**Theoretical implications of endowment effects**

Endowment effects also have some theoretical implications. The first one is that the Coase theorem (viz., in the absence of wealth effects and transaction costs, the final allocation is independent of initial entitlements) will not be valid if values are higher for holders of an entitlement than for potential buyers. Kahneman et al. (1990) randomly gave familiar consumption objects (pens and mugs) to one half of the subjects in an experiment. Subsequently, markets for these goods were conducted in which both
income effects and transaction costs were low. The observed volume of trade, which according to the Coase theorem, should be half of the mugs (assuming that half of subjects who like mugs were given a mug), was always significantly less.

The second theoretical implication of endowment effects is that indifference curves are no longer reversible, that is to say, the movement from A to B along the indifference curves is not the same as a movement from B to A (Knetsch, 1989). After all, if giving up a good has a greater impact on welfare than gaining the same entitlement, the rate of substitution at any point on an indifference curve is no longer the same for movements in either direction.

Finally, endowment effects imply the violation of completeness (Knetsch, 1995). According to the assumption of completeness, it is impossible to prefer good A over good B if it is framed as a loss, and at the same time prefer B over A if it is framed as a gain. For example, people are not willing to accept $10 as compensation (good B) when a forest is cut down (good A), meaning that they prefer A over B, whereas these people will not pay $10 (good B) to keep the forest (good A), meaning that they prefer B over A.

3.4.3 Overestimation

A third problem that is of great relevance for the CVM, is the persistent overestimation of true values.\textsuperscript{13} When discussing the study of Brookshire and Coursey in the previous section, we saw that the real payment (viz. the Smith auction process) gave significantly lower results than did the CVM field survey. This is only one example indicating that the CVM overestimates values when compared to more market-like elicitation methods. Apart from this study, many other studies have shown significant overestimations in CVM results. Here, only three of these studies will be reviewed.

Neill et al. (1994) tested the hypothesis of no significant difference between values reported by subjects, whether the payment of these values is real or hypothetical. They obtained data from controlled laboratory experiments that use two different market goods (an Indian painting and a 16th century map) and two valuation institutions (a CVM setting and a second-price Vickrey auction\textsuperscript{14}). In the painting experiment the mean WTP was $37.04 for the CVM (n=41) and $9.49 for the second-price auction (n=16), and the null hypothesis was rejected. The map experiment gave similar results.
Another study is described in Blackburn et al. (1994) and Cummings et al. (1995). Two experimental procedures were compared, using different market goods (electric juice extractors, chocolates, and in the second experiment also solar-powered calculators):

- Between subjects: one set of subjects is asked if they will buy a commodity at a given price (a 'yes' answer required a real payment), the second subset of subjects are asked the corresponding hypothetical question.
- Within subjects: subjects are asked the hypothetical question (“would you pay $x..” et cetera) and the good is then made available to the same subjects (“are you willing to pay $..” et cetera).

The results of the first experiment indicate that a significantly higher percentage of the subjects responded 'yes' to the hypothetical WTP question than to the real question: viz., 41% versus 16% for the juice machines and 42% versus 4% for the chocolates. This pattern is also found in the second experiment: viz., 41% versus 11% for the juice extractors, 42% versus 8% for the chocolates and 21% versus 8% for the calculators.

Thirdly, Bishop and Heberlein (1979) studied the WTA for duck-hunting permits. Depending on the bid format used, respondents in the hypothetical sample were willing to accept $68-$101 if their annual permit was taken away from them. A second group of respondents received a cheque which they could cash if they actually returned the permit. By varying the amount on the cheques, a mean 'true' WTA was established. The actual compensation demanded was considerably lower, namely $63. It was concluded that a hypothetical bias exists in the WTA format.

Explanations of overestimation
In order to explain the persistent and large overestimations, five interpretations are presented here:

1. hypothetical bias
2. strategic bias
3. question format
4. measurement of attitudes
5. lexicographic preference orderings
6. embedding

Hypothetical bias concerns the question whether the respondents' declared intentions (expressed WTP) can be taken as meaningful guides to their behaviour and as meaningful indications of their true values. True values are the values elicited in a real
market situation. For instance, when people buy potatoes, they have to ‘put their money where their mouth is’, that is, if the wrong choice is made, people will suffer real consequences. In a CVM questionnaire the same real-world implications are not present, and the risk of a hypothetical bias pertains if people will not put forth the same amount of effort in making a choice as the outcome does not affect them (Desvouges, 1995). Therefore, a misspecification of their true value results. For instance, Willis and Powe (1998) surveyed potential visitors to a castle prior to their entrance. They compared the maximum WTP of these potential visitors against the actual acceptance of an entrance ticket at their stated CVM price. Only 40% of the potential visitors at the castle’s gate were actually willing to buy the ticket at their stated price. Others refused to buy or declared they would visit the castle later. This evidence, as well as the evidence described above, suggests that most misspecification implies overestimation, not underestimation, of true values.

Another way to explain overestimations is by assuming strategic bias. Strategic bias occurs when a respondent gives a WTP amount that differs from his or her true WTP amount, in an attempt to influence the provision of the good and/or the respondent’s level of payment for the good. Strategic behaviour leads to overestimation if respondents believe they will not actually have to pay the amount they state, yet also believe that the stated amount can influence the provision of the good. Accordingly, they will overstate their true value in order to secure the provision. Milon (1989) studied strategic behaviour in CVM surveys. His results indicate that most strategic behaviour involves overestimation instead of underestimation. He tested three kinds of strategic behaviour using four different question formats:

1. strong free-riding, which means that people state a WTP ($WTP^s$) of zero while their true value ($WTP^t$) is positive
2. weak free-riding, denoting $WTP^s < WTP^t$ while $WTP^t > 0$
3. and overriding, implying $WTP^s > WTP^t$ while $WTP^t \geq 0$

Milon found that strong free-riding does not occur (the hypothesis that the $WTP^s$ is zero for the various question formats is rejected). To approximate the $WTP^t$, a closed-ended referendum mechanism is used with an equal per capita tax payment rule and majority vote implementation rule. Some results suggested weak free-riding, although there was more evidence of overriding. This finding is consonant with the conclusion of Mitchell and Carson (1989) in their state-of-the-art assessment of the CVM: “if strategic bias is present, slight overestimation is the expected result” (p. 168).
A third interpretation of overestimation deals with the dichotomous question format (e.g., Bennett and Tranter, 1998). When using this format, a respondent is asked whether or not she wants to pay, let us say, $100 for an increase in the provision of an environmental good. Suppose the respondent thinks that the interviewer will believe that she is not prepared to pay anything if she says 'no' to the $100 (the respondent reads the question as a choice between $0 and $100). Since respondents are inclined to display socially desirable behaviour, she says 'yes' to $100, whereas her true value could be something like $80. This is also known as compliance bias, where respondents overstate their true WTP in order not to appear to vote against environmental improvements.

Respondents might use the following decision rule in answering dichotomous questions: when asked the $100 question, they answer 'yes' if their true WTP is higher than $50 and 'no' if their true WTP is smaller than $50. The use of these and similar decision rules leads to severe overestimation. Several CVM studies support the interpretation of overestimation due to the use of dichotomous question formats. E.g., Boyle et al. (1996 and 1998), Brown et al. (1996), Halvorsen and Sælensminde (1998) and Ready et al. (1996) compared different question formats, and all of these studies demonstrated that the dichotomous format yields (much) higher answers than the open-ended format, especially for hypothetical payments as in the case of the CVM.

A fourth explanation of overestimation is the fact that people might state an overall preference, attitude, intention, or ideal towards the environment, rather than a well-structured preference. Earlier (chapter 2, section 2.1.2.f), this tendency was referred to as warm glow, moral satisfaction, good cause dump, or symbolic bias. In short, such an overall preference involves a higher value than would the preference for a more specific good, and thus entails overestimation.

A fifth explanation of overestimation is that individuals' ethical attitudes could give rise lexicographic preference orderings (cf. chapter 2, section 2.1.2.a) of ordinary goods and environmental goods, based on a moral commitment in favour of environmental protection. For instance, species preservation is always preferred to more income as long as income is above some threshold level (Perman et al., 1999, pp. 415-416).

Finally, the cause of overestimation in CVM results might be the embedding effects, which were discussed in section 3.4.1. If a good is valued on its own, the estimated value will be much higher than if the good is valued as part of a more inclusive good. Embedding effects appear to be an important cause of overestimation (Kemp and Maxwell, 1993).
'Solutions' to overestimation

The problem of overestimation has been 'dealt with' by the NOAA panel. As was mentioned earlier, this panel was formed for the purpose of considering whether or not one can derive CVM values that are reliable and valid. The panel's recommendations are sometimes treated as laws for CVM studies although these recommendations have not yet been thoroughly tested for their merits (Schulze et al., 1996). One such law is the following calibration rule, which is an attempt to deal with overestimation. In their 1993 report, the panel states that it is

"persuaded that hypothetical markets tend to overstate willingness to pay for private as well as public goods. The same bias must be expected to occur in CVM studies." (Arrow et al., 1993, p. 4610)

In a more recent report the panel proposes a rule to correct for this hypothetical bias.

"The proposed rule provides that the respondents' stated values be divided by two [emphasis added, BB]. This calibration procedure is to 'correct' for the combined effects of two countervailing potential biases: the mandated elicitation of willingness to pay (WTP) measures may understate the correct measure of damages (the willingness to accept), whereas the elicitation of hypothetical WTP in contingent valuation studies may overestimate 'true' WTP." (Arrow et al., 1994, p. 12)

As was noted earlier (section 3.4.2), this is a rather inexact, unscientific rule. For example, calibration by dividing CVM responses by two assumes the existence of monotonic links between CVM responses and true values (Randall, 1996). Furthermore, the value of two is disputable. Fox et al. (1998) tested a calibration procedure using the WTP of consumers for risk reduction of food irradiation. Their results suggest, as common sense also predicts, that the calibration factor is good-specific and context-specific. In other words, there is no generally valid value for a calibration factor for hypothetical results.

Others are harsher in their critique of the calibration rule, e.g., Knetsch (1994) who refers to this 'solution' as a nearly comical response to CVM failures and inappropriate selection of measure (WTP instead of WTA). Moreover, these critics argue that the proposal would fail to hold polluters fully accountable for the damage they cause and would not provide a strong incentive for prevention.
The question is where the 'two' in 'divided by two' comes from. The story goes that, some 12 years ago, a researcher, discussing a forestry study, cried out: “But these values are much too high; let’s divide them by two". Some other researchers asked him why the values should be divided by two, and not by three or some other value. He replied: “Well, because of the following divine argument: half of the forest was created by the Forestry Service and the other half was created by God and therefore should not be valued". Of course this is just a joke, but it does make clear the scientific controversy about how to deal with overestimation.

3.4.4 Aggregation Issues

The purpose of a CVM study is to provide decision makers or judges with the valuation of a certain change in the provision of an environmental good. To obtain such a measure of total benefits and costs, that is of social welfare, the individual welfare measures must be added up over the relevant population. This raises the question of how the welfare of individuals is related to social welfare.

This question has been dealt with by Arrow (1963). Arrow’s impossibility theorem states that, given some minimal and separately reasonable requirements, it is impossible to devise a method to aggregate individual preference orderings to a consistent collective preference ordering. To derive this result, he assumed that the only information available is the preference ordering of the individuals (e.g., utility functions are ordinal), and that the decision maker is precluded from making interpersonal comparisons of welfare. However, by adding more information on individual utilities (individual utilities can be made measurable, e.g., cardinal) and by invoking somewhat stronger value judgements (utilities can be made comparable between individuals), it is possible to get a consistent collective preference ordering (Hennipman, 1995). Below, the concept of ordinality versus cardinality and of interpersonal comparability versus non-comparability will be reviewed first. After that, aggregation issues are discussed.

Ordinality versus cardinality

The controversy between the ordinal and the cardinal measurability of utility began when Vilfredo Pareto (1858-1923) opposed against the notion of cardinality. Pareto introduced the idea that ordinal measurability of utility is sufficient to effectively describe consumer behaviour. He argued (in: Van den Doel and Van Velthoven, 1990, p. 36) that:
"A man can know that the third glass of wine gives him less pleasure than the second; but he can in no way tell what quantity of wine he must drink after the second glass in order to get pleasure equal to that which the second glass provided him."

Ordinality implies that the order, not the absolute value, of the numbers is relevant. In terms of utility this implies that marginal utility has meaning only in being positive or negative, but that the numerical value is meaningless (Alchian, 1953). Such an ordinal utility function is not unique, since various positive monotonic transformations \( \psi \) of the personal welfare function \( W_i \) are possible by which the order remains the same for all individuals: \( W_i^* = \psi(W_i) \). In the standard welfare economic analysis, it is generally assumed that \( W_i \) is not interpersonally comparable between individuals (Roberts, 1980, pp. 423; Sen, 1977, pp. 1541-1542).

A utility function is cardinal when utility differences can be compared for one individual as well (intrapersonal comparison). A cardinal function is unique up to a positive affine transformation, i.e., if \( W_i \) is a cardinal function, \( W_i^* = \alpha_i + \beta W_i \) and \( \beta > 0 \) elicits the same information (ibidem). If it is also assumed that preferences are interpersonally comparable the possibilities of informed welfare judgement are widened (Sen, 1979, p. 546).

However, cardinal utility measurement with interpersonal comparisons is a reasonable assumption within a relatively homogeneous group of people, that is homogeneous with respect to income, cultural characteristics, and social norms and values. Moreover, without these assumptions it would not be possible to empirically study changes in the level of utility (i.e., welfare or well-being).

The valuation methods discussed in this thesis are based on interpersonal comparability (among other things, when aggregating values across individuals). For instance, Van Praag (1971, welfare evaluation) directly asks people to evaluate their income, i.e., subjects are asked to indicate which income level they consider 'very bad', 'bad', 'insufficient', 'sufficient', 'good' and 'very good'. Because it is assumed that the verbal levels have the same meaning to all respondents, interpersonal comparisons are possible. Similarly, Cantril (1965, well-being evaluation) directly asks people to evaluate their life as a whole (well-being) on a ladder of life with 10 steps. It is assumed that people positioning themselves on the same step, experience the same level of well-being. These methods will be discussed in greater detail in chapter 4.
Aggregation issues

Measuring welfare in the cardinal sense has been rejected by economists since the late 1930s. Instead, an ordinal concept of utility was adopted, with no comparability across individuals. One welfare criterion that complies with the ordinal utility concept, is the so-called Pareto criterion. The Pareto criterion is a weaker welfare measure than the cardinal principle (i.e., maximizing the sum of personal welfares, Sen, 1977, p. 1546), in the sense that it does not facilitate the identification of a unique welfare optimum for society. The criterion is based on examining all possible incremental reallocations of resources among individuals and continuing to accept such changes so long as they improve the well-being of at least one individual without harming that of others. Although most economists reject cardinality as well as interpersonal comparability, they use it all the time. For instance, the national income is calculated by simply adding total expenditures across consumers and investments across firms. Another example is the fact that firms and government agencies weigh up the costs and benefits of their investments or policies. The Pareto criterion is seldom used in practice.

When aggregating individual money values economists use the idea of ‘one dollar, one vote’, which implies that the current income distribution is taken as given and only efficiency is dealt with. In the CVM, such aggregations could be interpreted as saying that the people with the largest budgets (the rich) care most about the environment. After all, the expressed WTP depends not only on preference but also on the ability to pay.

One method to correct for large income disparities is by weighting. If everyone had an equal income, a bid of f10 from John would mean the same thing as a bid of f10 from Mary. But if Mary has a million guilders and John has only f10,000, under those circumstances, to count Mary’s bid as meaning the same as John’s is probably ludicrous, given the fact that the marginal utility of money varies with income. The income inequality can be adjusted by giving John’s WTP a weight of, let us say, 50 (assuming a decreasing marginal utility of income) and Mary’s bid a weight of one.

The United Nations Environmental Programme (UNEP, 1994) suggests that the weighting of income distribution is an issue to be addressed in the case of using CVM studies in developing countries, because the costs and benefits are different in kind and also accrue to different people, as absolute levels of income are lower in developing countries and often greater disparities exist between rich and poor. Nevertheless, in general income weighting is not used. One of the reasons might be the fact that, to be consistent, income distribution should be reflected in all decisions.
(not just in the decision evaluated in a CVM study), and should, moreover, concern not only intragenerational justice but also intergenerational justice (Pearce, 1993).

However, it is not very sensible for a researcher to use income weighting in his or her research when conducted within a certain society or in similar societies, since in the real world (the market) high incomes are also related to higher abilities to pay for something. On the other hand, in the case of income differences between very different societies, income weighting is necessary to arrive at sound conclusions. If you do not correct for these differences in income between very different countries, results can occur that are politically unacceptable, as the following example demonstrates.\(^\text{19}\)

The implicit conclusion of a global CVM study by the International Panel on Climate Control (IPCC), was that the value of a human life in a poor country was $250,000, while that value was $2,500,000 in the industrialised world. The study was done in order to provide guidelines to the global community about the issue of global warming and the kinds of projects that are to be implemented in order to deal with the resulting deaths. The study was designed and carried out by some respected economists like Nordhaus. They concluded that a project whose costs are, say $25,000,000, should be undertaken in a developed economy if it saves 10 lives, but that this same project must not be implemented in a poor country unless it saves 100 lives. This result was, of course, rejected by several less developed countries, including India, China, Cuba, Brazil and Peru, at a meeting of the IPCC on July 28, 1995. The economists were told to go back and do their work over again. Since then, a lively discussion has ensued, but there is still no consensus on the aggregation of statistical lives across countries.

### 3.5 Concluding Remarks

"Among economists there is an old belief which, in an extreme form, makes them feel convinced of having the truth. Others are willing to settle for less, but, even in the opposite extreme, in their wavering minds they have the shimmering idea that they are on a very special road. Which of course is true. This belief, of which no one needs to be ashamed, is the belief in the 'scientific nature' of economics." (Klant, 1987, p. 33)\(^\text{20}\)

This quote exemplifies the search for the truth, or for true values, that is also apparent in the CVM. One of the principal assumptions underlying the CVM is that people have true, but hidden, economic values for environmental goods which can be revealed through the creation of a hypothetical market. The question is whether these true
values exist, or whether several true values exist. Several true values might exist if existing costs and benefits are the product of a set of property rights, the economic system, income distribution and so on. If you change the starting point, a different value will result (Schmid, 1995). A second question is whether the CVM can elicit true values. CVM believers assume that, if procedural biases are absent, a neutral survey could convert subjective feelings into scientifically viable expressions of value (Harris et al., 1989). For example, the NOAA panel concluded that, if conducted under appropriate conditions specified by the panel,

"CVM studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values." (Arrow et al., 1993, p. 4610)

Although the CVM pretends to obtain these scientific estimates (since they are related to theoretical concepts like compensating variation and equivalent variation), I do not think that the CVM—or any other stated preference method for that matter—is capable of eliciting an unambiguous, theoretically correct value. The WTP principle on which such estimates are (supposedly) based, is elastic enough to allow a huge range of estimates. Apart from anything else, it shows that different assumptions and procedures will lead to widely different results.

For instance, the implication of several biases, framing effects, embedding effects and endowment effects is that the CVM is open to political influence: by framing the questions in a certain way, by embedding the good in a larger context or by using a WTP measure instead of a WTA measure, CVM practitioners can deliberately influence the results so as to please themselves or their sponsors. In other words, the method appears to be susceptible to serious manipulation, which is not a good feature for a method on which damage awards and allocation decisions are to be based and which claims to be scientific (unambiguous). Some people even suggest that it seems to be a case of 'tell me the figure you'd like, and I'll provide a justification' (Diamond and Hausman, 1994).

This latter remark goes too far in its sarcasm towards the CVM. In short, the main benefit of the CVM and other monetary valuation methods lies in the discipline it imposes, not in the bottom line that creative practitioners are able to squeeze out of it. The CVM is not science 'pur sang', in the sense that it measures unambiguous or true values, but rather a systematic procedure for collecting and organizing information that can be used to make decisions. It is a decision tool. So, instead of using the CVM as an analytical tool to provide scientific values, the CVM should be used as a policy
decision method. In that case, the primary purpose of the CVM is not to consider what the true price of a particular environmental good would be, but to estimate the subjective values for the good in question.

As was stated at the outset of this book (chapter 1), this thesis has a twofold objective. The first is to critically review the CVM and its theoretical basis and assumptions. So far we have focused mainly on this first objective. We have seen that, despite the problems and criticisms related to the CVM, it is the most well-known and most popular valuation method. Now it is time to turn to the second objective, namely to introduce and test alternative methods to the CVM.
Endnotes

1 Hoevenagel (1994, pp. 8-10) also distinguishes non-preference methods, which are not included in the classification presented here.

2 Revealed preference methods are based on observable market behaviour as represented by Marshallian demand curves. The Marshallian demand curves track the price effects which occur when the provision of a commodity changes. Revealed preference methods measure the income changes necessary to neutralise these price effects, the so-called consumer surpluses. Hickian demand curves, on the other hand, are not observable since they depend on utility. Therefore, these curves can only be estimated when using stated preference methods. Stated preference methods measure the income necessary to neutralise changes in utility caused by changes in the provision of some environmental good, the so-called compensating and equivalent variations (cf. chapter 2, section 2.1.1).

3 It is important to note that CVM measures provide estimates of total values, whereas surrogate market approaches (like the travel costs method) only provide estimates of use values (cf. section 3.3).

4 Carson et al. (1992) interviewed 1,043 people across the US and asked how much they would pay in a one-time tax for a plan to provide two Coast Guard ships and trained personnel to escort oil tankers in the Prince William Sound, to prevent future accidents and avoid future injuries due to oil spill. They found a mean WTP of $31, which makes $2.8 billion when multiplied by the number of American households.

5 Perfect embedding implies regular embedding.

6 Temporal embedding does not have to be a problem for the CVM, since it also arises in real markets. For instance, when you buy a television or a car on hire-purchase, the total sum of the periodic payments will be more than when you pay the purchase price at once, even if you account for interest.


8 Mental accounts can be related to the theory of two-stage budgeting (Deaton and Muellbauer, 1980), in which total expenditure is first allocated to broad categories, such as food, clothing and housing, and then each allocation is divided among specific items in each category.

9 Hanemann's basic result is: price flexibility of income = $\mu / \sigma_0$, where $\mu = 0$ means there are no income effects (zero income elasticity for the environmental good) and $\sigma_0 = \kappa$ means that the environmental good and the other good(s) are perfect substitutes (p. 641).

10 The SAL procedure is more similar to market situations because the payments are not hypothetical (the fund consisted of money that was actually collected or distributed from or to the group), and because respondents could revise their bids (up to five times). If the group either did not cover the costs of an expansion of trees (WTP) or did not request more compensation than available for a reduction in the number of trees (WTA), another trial was conducted. Individuals could adjust their bids in between these trials.

11 Since value is defined on deviations from the reference point, the WTA/WTP disparity is sometimes referred to as 'status quo bias'. Similarly, since the value function is concave for losses and convex for gains, the disparity is sometimes labeled 'loss aversion' (Kahneman et al., 1991).

12 The asymmetry between losses and gains is even recognised by US court decisions "by favoring possessors of goods over other claimants, by limiting recovery of lost profits relative to compensation for actual expenditures and by failing to enforce gratuitous promises that are coded as forgone gains to the injured party" (Kahneman et al., 1990, p. 1346).
However, overstatement is not a typical CVM problem; it can occur in all surveys.

This is a valuation institution in which truth-telling is a dominant strategy, since the highest bidder will have to pay the second highest bid.

This story was told to me by George Peterson at the Joint European Conference on Non-Market Valuation in Oslo, in June of 1995.

These requirements include: the condition of unrestricted domain, the Pareto rule, the condition of independence of irrelevant alternatives, and the absence of a dictator (Arrow, 1963, pp. 22-31).

Until that time most economists, from Smith and Bentham in the 18th and the beginning of the 19th century, to Jevons and Walras in the 19th century and Pigou in the 1920s, believed that utility was a cardinal quantity.

The Pareto criterion is too restrictive to be of much relevance in a policy context, as it is virtually impossible to imagine a public project that would not impose net costs on someone in society. Independently, Kaldor and Hicks both proposed two slightly different forms of a more operational welfare criterion known as the potential compensation test or potential Pareto improvement. The central feature of the Kaldor version of this test is whether those who gain from a policy change could, in theory, compensate losers in such a way that at least one individual would be better off and no one would be worse off. Compensation needs to be only hypothetical; the potential Pareto improvement does not actually have to be implemented.

These results were presented on the ecol-econ internet discussion network in June of 1995 (ecol-econ@csf.colorado.edu).

This quote is originally in Dutch. Translation by Ada Kromhout.

The term 'true values' refers to the compensating and equivalent variation measures that would be elicited if real markets existed for environmental goods.