

5.7 Anwendungsprobleme der Contingent Valuation Method

In der Literatur werden verschiedene Anwendungsprobleme der Contingent Valuation Method diskutiert, die zu Verzerrungen bei den Ergebnissen führen können und somit die Güte der Ergebnisse beeinträchtigen. Verzerrungen in diesem Sinn liegen dann vor, wenn davon ausgegangen werden kann, daß die mit Hilfe der Contingent Valuation ermittelten Werte nicht denen entsprechen, die die Individuen auf realen, unverzerrten Märkten äußern würden. Aus Sicht der Kritiker ist daher fraglich, ob die von Befragten gegebenen Antworten tatsächlich deren maximale Zahlungsbereitschaft repräsentieren und wie stark geäußerte und wahre Zahlungsbereitschaft voneinander abweichen. ELSASSER (1996: 75) folgend läßt sich die Kritik an den Anwendungsproblemen der CVM in zwei Linien unterteilen. Der eine, in seiner Kritik gemäßigte Ansatz geht davon aus, daß die geäußerte Zahlungsbereitschaft zwar auch von den wahren Präferenzen der Nutzer geprägt würde, darüber hinaus aber von anderen, situationsspezifischen Einflüssen verfälscht würden. Diese Position kann als Hypothese schwacher „Verzerrungen“ bezeichnet werden.

Darüber hinaus gibt es Kritiker, die dem Ansatz der „starken Verzerrung“ zuzuordnen sind. Ihnen folgend wäre die geäußerte Zahlungsbereitschaft von der wahren Zahlungsbereitschaft der Befragten völlig unabhängig. Sie stelle lediglich eine Funktion anderer Einflüsse dar. Zu den *anderen Einflüssen* wird vor allem der hypothetische Charakter der Contingent Valuation Method gezählt, an dem dann weitere Verzerrungsmöglichkeiten festgemacht werden. Die Aufzählung gibt einen Überblick über die Verzerrungsmöglichkeiten, die intensiv diskutiert wurden bzw. noch werden:

- Hypothetischer Charakter (Hypothetical Bias),
- Strategisches Verhalten (Strategic Bias),
- Einfluß des Zahlungsinstrumentes (Payment Vehicle Bias),
- Startwertverzerrung (Starting Point Bias),
- Zuordnungsfehler (Part-Whole-Bias) und
- Beeinflussung durch Informationen.

5.7.1 Gütekriterien für die Beurteilung der Ergebnisse der Contingent Valuation

Bevor auf die oben genannten Verzerrungsmöglichkeiten näher eingegangen wird, sollen noch zwei Kriterien zur Beurteilung der Güte der ermittelten empirischen Daten

vorgestellt werden: die Reliabilität und die Validität.¹⁶ Den Anforderungen der *Reliabilität* wird dann entsprochen, wenn gezeigt werden kann, daß die Antworten der Befragten nicht beliebig waren. Dies kann durch eine spätere Wiederholung der Befragung, durchaus auch mit einem kleineren Teil der ursprünglichen Stichprobe, überprüft werden. Da dieses Verfahren aber sehr kostenintensiv ist, wird in der Regel eine Regressionsanalyse zur Überprüfung der Reliabilität eingesetzt. Dabei wird angenommen, daß je höher der Anteil der durch die unabhängigen Variablen erklärten Varianz an der Gesamtvarianz einer Schätzungsgleichung für die Zahlungsbereitschaften ist, desto geringer ist der restliche, nicht oder nur durch Zufall erklärbare Anteil an den Antworten.

Das Bestimmtheitsmaß¹⁷ R^2 kann als das einfachste Maß zur Beurteilung des Reliabilität angesehen werden. Es gibt jedoch nur dann einen guten Eindruck der Reliabilität, wenn es aus Einflußvariablen gewonnen wurde, die plausiblerweise einen relevanten Einfluß auf die Zahlungsbereitschaft haben. Für R^2 muß ein bestimmter Funktionszusammenhang unterstellt werden (meist linear oder loglinear). Nur wenn dieser angenommene Zusammenhang die Verhältnisse in der Realität trifft, dann ist R^2 ein gutes Maß für den Erklärungsgehalt einer Regression. CARSON und MITCHELL (1989) fordern, daß mit wenigen zentralen Variablen ein R^2 von mindestens 0,15 erreicht werden sollte.

Mit dem Kriterium der Validität soll überprüft werden, ob die Resultate mit theoretischen Überlegungen oder Nutzeneinschätzungen anderer Verfahren konsistent sind. Der theoretischen Validität entspricht ein Verfahren, wenn es mit den aus der Nutzentheorie eingeleiteten Hypothesen übereinstimmt. Auch die theoretische Validität kann mit Hilfe der Regressionsanalyse untersucht werden. Hierzu werden jene erklärenden Variablen (z. B. Einkommen, Bildung, Alter) herangezogen, deren Einfluß auf die Zahlungsbereitschaft aus dem theoretischen Modell abgeleitet werden kann.

Tabelle 7: Zusammenfassung zu den Kriterien der Reliabilität und Validität

Kriterium	Definition
Reliabilität	<ul style="list-style-type: none"> • Maß für den Erklärungsgehalt, den die zugrunde gelegten Einflußvariablen (z. B. Einkommen, Bildung, Alter) auf die geäußerte Zahlungsbereitschaft haben. Eine Untersuchung hat einen hohen

¹⁶ Vgl. u.a. MITCHELL, CARSON (1989): 189ff; BISHOP ET AL. (1995); BISHOP ET AL. (1997); GRONEMANN, HAMPICKE (1997) sowie ENDRES, HOLM-MÜLLER (1998).

¹⁷ Das Bestimmtheitsmaß ist eine Maßzahl zur Beurteilung der Güte der Anpassung eines Regressionsmodells. Es bietet eine Entscheidungshilfe darüber, ob die exogenen Variablen insgesamt die endogenen Variablen erklären können (vgl. ECKEY ET AL. 1995: 50).

	Grad an Reliabilität, wenn der Grad der unerklärten Varianz niedrig ist.
Validität	<ul style="list-style-type: none"> • Läßt sich am treffendsten mit Verlässlichkeit übersetzen; Validität mißt systematische Verzerrungen zwischen dem theoretischen Konstrukt und der erhobenen Meßgröße. Konzept der Validität ist in drei Dimensionen unterteilt: <ul style="list-style-type: none"> → <i>Inhaltliche Validität (content validity)</i>: bezieht sich in erster Linie auf die Identität des tatsächlichen Objektes der Messung mit dem intendierten Objekt → <i>Kriterium-Validität (criterion validity)</i>: Frage, ob eine verwendete Meßgröße überhaupt prinzipiell geeignet ist, den Untersuchungszustand zu erfassen; d.h. in diesem Fall die Zahlungsbereitschaft; Problem ist hier vor allem die hypothetische Situation, da in realen Märkten auch gezahlt werden muß → <i>Konstrukt-Validität (construct validity)</i>: Frage, ob sich die Meßgröße (die erfragte Zahlungsbereitschaft) zu andern Größen so verhält, wie es die Theorie voraussagt • <i>Konvergenz Validität (convergent Validity)</i>: Mißt die Korrelation der Meßgröße mit anderen Meßgrößen für dasselbe theoretische Konstrukt. • <i>Theoretische Validität (theoretical validity)</i>: Stellt die Frage, ob sich die Meßgröße zu anderen Größen so verhält, wie es aus theoretischen Überlegungen heraus erwartet wird.

Quelle: zusammengestellt u.a. nach ENDRES, HOLM-MÜLLER (1998: 74ff)

5.7.2 Strategisches Verhalten (*Strategic Bias*)

Mit dem strategischen Verhalten wird eine Möglichkeit zur Verzerrung angesprochen, die auf das in der Ökonomie nicht nur in bezug auf die Zahlungsbereitschaftsanalyse viel diskutierte Problem des *Trittbrettfahrers* - oder auch *free riders* – zurückzuführen ist. Der Trittbrettfahrer zeichnet sich dadurch aus, das er nicht seine tatsächlichen Präferenzen offenbart sondern sich durch "falsche" Angaben einen individuellen Vorteil verschaffen will. Zwei mögliche Fälle sind zu unterscheiden.

- Der Befragte geht davon aus, daß er zur Zahlung für die Bereitstellung des öffentlichen Gutes herangezogen wird. Da er selbst seinen Einfluß auf die gesamte Nachfrage für unwesentlich hält, nennt er einen Betrag, der unterhalb seiner eigentlichen Präferenzen liegt. Dabei geht er davon aus, daß durch die Präferenzäußerungen der anderen und deren Zahlungen das öffentliche Gut in einer auch für ihn selbst ausreichenden Menge bereitgestellt wird. Damit käme der Trittbrettfahrer in den Genuß des öffentlichen Gutes ohne aber entsprechend seiner tatsächlichen Präferenzen dafür zahlen zu müssen.
- Der Befragte nimmt umgekehrt an, daß er nicht zur Zahlung eines Beitrages herangezogen wird. Vielmehr geht er davon aus, daß die Bereitstellung des öffentlichen

Gutes durch das allgemeine Budget finanziert wird. In dieser Situation besteht nun ein Anreiz, eine höhere Wertschätzung als tatsächlich vorhanden anzugeben.

Strategisches Verhalten in diesem Sinn wird als eines der Hauptprobleme der direkten Bewertungsmethode angesehen. Die Ergebnisse empirischer Untersuchungen legen jedoch nahe, daß dies aus der Theorie abgeleitete Argument überbewertet wird. Es konnte gezeigt werden, daß die Befragten sich nicht in dem erwarteten Maß strategisch verhalten haben (vgl. POMMEREHNE, SCHNEIDER 1980, RÖMER 1991, WEIMANN 1996b). Dies soll umgekehrt nicht heißen, das die Bedeutung strategischen Verhaltens bei der Durchführung von CVM-Studien zu vernachlässigen ist. Insbesondere beim Aufbau von Fragebögen muß dies berücksichtigt werden. So empfehlen z. B. BATEMAN und TURNER (1993), den Befragten zusammen mit der Frage nach ihrer Zahlungsbereitschaft zu vermitteln, daß bei insgesamt zu geringer Zahlungsbereitschaft das betreffende Produkt auch nicht angeboten bzw. erstellt würde. Umgekehrt würde die Information, daß die Kosten für die Produkterstellung letztlich vom Staat, von der Allgemeinheit getragen würden, einen starken Anreiz zu strategischem Verhalten bieten. POE et al. (1997) berichten von Ergebnissen empirischer Untersuchungen, die die positive Anreizwirkung sog. Zahlungs- und Implementierungsregeln¹⁸ bestätigen. Sie haben die Ergebnisse einer Umfrage für die Zahlungsbereitschaft zur Förderung regenerativer Energieträger, in denen den Befragten diese Regeln präsentiert wurden, mit den Ergebnissen eines tatsächlichen Marktes, den sie parallel zur Befragung errichtet haben, verglichen und eine hohe Übereinstimmung zwischen den Ergebnissen festgestellt. Dies sehen sie als Beleg dafür an, daß auch hypothetische Märkte anreizkompatibel ausgestaltet werden können. In Deutschland hat RÖMER (1993) vergleichbare Zahlungs- und Implementierungsregeln eingesetzt, allerdings nicht explizit ausgewertet, wie sie sich auf die geäußerten Zahlungsbereitschaften ausgewirkt haben.

5.7.3 Hypothetischer Charakter (*Hypothetical Bias*)

Der hypothetische Charakter der Befragung kann dazu führen, daß die Individuen - zwar eher unbewußt und nicht als Resultat strategischen Verhaltens - systematisch von ihren tatsächlichen Wertschätzungen abweichen. Dies wird einmal darauf zurückgeführt, daß die Korrelation zwischen Einstellung und Verhalten zu niedrig sei, um

¹⁸ Poe et al. (1997) haben sinngemäß folgende drei Regeln verwendet:

1. Kommt dabei genug Geld zusammen, um die Kosten für das Programm zu decken, dann würden die Maßnahmen durchgeführt.
2. Sollte mehr Geld als notwendig zusammenkommen, wird die Maßnahme natürlich auch durchgeführt. Das übrige Geld würde aber an die Haushalte entsprechend zurückgezahlt.
3. Wenn jedoch weniger Geld zusammenkommt, dann können die Maßnahmen nicht durchgeführt werden. In diesem Fall bekommen die Haushalte ihr gesamtes Geld zurückgezahlt.

aus einer Frage nach einem möglichen Verhalten Rückschlüsse auf späteres wirkliches Verhalten abzuleiten. Zum anderen würde die hypothetisch geäußerte Bewertung von der wahren Wertschätzung abweichen, da es in der hypothetischen Situation nicht genügend Anreize gebe, die eigene Wertschätzung durch sorgfältiges Abwägen festzulegen. RÖMER (1991: 422 f.) merkt dazu folgendes an:

- Zwischen Einstellung und Verhalten ist fast immer dann eine Korrelation nachweisbar, wenn die Befragten in möglichst konkrete Situationen versetzt werden. So müßten z.B. Ziel, Handlungsalternativen, Kontext und Zeitrahmen der Frage genau festgelegt werden. Dieser Art des Repräsentanzproblems kann durch eine geeignete Ausgestaltung der Befragung begegnet werden.
- Laborexperimente und entsprechend angelegte Untersuchungen würden zeigen, daß weder Anhaltspunkte für ein deutliches Abweichen noch für eine gänzliche Übereinstimmung von hypothetisch-kontingenter und experimentell-simulierter Bewertung vorliegen.
- Schließlich sei festzuhalten, daß der hypothetische Charakter auch positive Folgen haben könne, da die Anreize zum strategischen Verhalten deutlich geringer sein können.

5.7.4 Einfluß verschiedener Zahlungsinstrumente (*Vehicle Bias*)

Die Angabe des Zahlungsinstrumentes ist ein wesentlicher Bestandteil der Konstituierung des hypothetischen Marktes. Denkbar sind als Zahlungsinstrumente z. B. eine einmalige Abgabe, die Erhöhung der Einkommenssteuer oder einer indirekten Steuer wie der Mineralölsteuer, die Erhebung eines Eintrittsgeldes oder Spenden an einen Fond zur Erhaltung der Flußauen.

Entsprechend ausgerichtete Studien haben gezeigt, daß die Ausgestaltung des Zahlungsinstrumentes Einfluß auf die Höhe der Zahlungsbereitschaft hat bzw. im Extremfall sogar zu Protestantworten führen kann. Letzteres kann z. B. eintreten, wenn als Zahlungsinstrument eine Erhöhung der Steuer angegeben wird, nach Meinung der Befragten die Steuerlast schon ohne diese weitere Erhöhung zu hoch ist und dem Staat auch kein Vertrauen entgegen gebracht wird, daß er die Mittel entsprechend der genannten Zielsetzung verwendet, sondern die erhobenen Beiträge im allgemeinen Haushalt verschwinden würden. So kam HAMPICKE z. B. im Rahmen einer Untersuchung über die Zahlungsbereitschaft für den Arten- und Biotopschutz in der BRD zu dem Ergebnis, daß an staatliche Stellen wesentlich weniger gezahlt würde als an unabhängige Stellen (HAMPICKE 1991: 133). Andererseits kann eine Steuer aber auch positive Auswirkungen auf die Zahlungsbereitschaft haben, da die Befragten dann davon ausgehen, daß im Falle der Bereitstellung des öffentlichen Gutes auch alle ande-

ren zur Finanzierung herangezogen werden und dadurch Trittbrettfahren ausgeschlossen wird.

Als Konvention, so ELSASSER (1996: 25), habe sich daher herauskristalliert, daß in der jeweiligen Situation plausibelste Zahlungsinstrument vorzustellen. Daher empfiehlt es sich, verschiedene Instrumente in Pretests auf ihre Akzeptanz hin zu untersuchen, um hieraus möglicherweise resultierende negative Beeinflussungen der eigentlichen Untersuchung so weit möglich auszuschließen (vgl. auch HANLEY et al. 1996: 20). Bei der Wahl des Zahlungsinstrumentes zu berücksichtigen ist, daß sich nicht alle Zahlungsinstrumente eignen, wenn Non-use Values (mit) ermittelt werden sollen. So können diese z. B. nicht über Eintrittspreise ermittelt werden, zumindest nicht bei Personen, deren Wertschätzung ausschließlich aus der Nicht-Nutzung resultiert.

5.7.5 Zuordnungsfehler (*Part-Whole Bias / Embedding- Effect*)

Die Diskussion um den Part-Whole Bias¹⁹ bezieht sich im wesentlichen auf einen Artikel von KAHNEMAN und KNETSCH aus dem Jahr 1992. Darin führten sie anhand der Ergebnisse einer eigenen Umfrage aus, daß die Validität der Contingent Valuation Method nach ihrer Einschätzung fraglich sei: Zum einen würde ein Embedding-Effekt auftreten, da die Bewertung des (Umwelt-) Gutes davon abhängig sei, ob es als eigenständiges Gut oder als Teil eines umfassenderen Güterbündels präsentiert würde. Zum anderen würde die Reihenfolge, in der die Präsentation der Güter erfolgen würde, entscheidenden Einfluß auf die WTP haben. D.h., die jeweils zuerst genannten Güter würden die höchste Zahlungsbereitschaft auf sich ziehen, während nachfolgend genannte nur noch eine geringere Zahlungsbereitschaft auf sich ziehen könnten. Diesen zweiten Effekt erklären KAHNEMAN und KNETSCH mit der „moralischen Befriedigung“, die die Befragten aus ihrem Beitrag für eine gute Sache ziehen würden (auch als *Warm Glow of Giving* bezeichnet). Es bestehe keine Präferenz für das jeweils zu bewertende Gut, sondern eine für das Gut „moralische Befriedigung“, also etwas Gutes zu tun. Daher könne aus den mit Hilfe der Contingent Valuation gewonnenen Informationen auch nicht auf den ökonomischen Wert, d. h. die hierfür jeweils bestehende Wertschätzung auf Grundlage der individuellen Präferenzen, geschlossen werden. Vielmehr komme lediglich eine moralische Befriedigung zum Ausdruck.

Vor dem Hintergrund dieser Kritik wird daher als Anforderung an die Contingent Valuation Method formuliert, daß sie für Situationen, für die die ökonomische Theorie

¹⁹ Für den Part-Whole-Bias werden, wie schon Überschrift und erste Zeilen des Absatzes zeigen, unterschiedliche Begriffe verwendet. In der englischen und amerikanischen Literatur wird häufig auch als Oberbegriff der Terminus „Scope“ (Scope Insensitivity) verwendet (z. B. CARSON 1997). Von ELSASSER wurde der Begriff des Zuordnungsfehlers als Übersetzung eingeführt. In anderen Studien findet aber auch der Embedding-Effekt als Oberbegriff Verwendung (z. B. DEGENHARDT, GRONEMANN 1998).

unterschiedliche Werte vorhersagt bzw. erwarten läßt, auch tatsächlich verschiedene Ergebnisse liefert. Würden die von KAHNEMAN und KNETSCH erhobenen Vorwürfe gegenüber der CVM in Gänze zutreffen, dann wäre die Methode als Instrument zur Ermittlung ökonomischer Werte weitgehend diskreditiert und die damit gewonnenen Werte könnten weder als Grundlage für die Festlegung von Schäden in Gerichtsverfahren noch für Kosten-Nutzen-Analysen herangezogen werden. Dementsprechend hat sich im Anschluß an diese Veröffentlichung eine sehr umfassende und intensive Diskussion um die Existenz des Embedding-Effektes und des Warm-Glow-Effektes - beide werden als den Part-Whole Bias konstituierend angesehen - entwickelt.

Zur Illustration soll folgendes Beispiel dienen (vgl. Tabelle 8): Aufgrund der ökonomischen Theorie wäre zu erwarten, daß die Zahlungsbereitschaft für das Gut „Erhalt der Flußauen in Deutschland“ höher sein müßte als für das Gut „Erhalt der Flußauen in den neuen Bundesländern“ als wiederum für das Gut „Erhalt der Flußauen an der Saale“. Würden verschiedene Gruppen nach jeweils allen drei Gütern, nach den beiden letzten oder lediglich nach ihrer Wertschätzung für das Gut „Erhalt der Flußauen an der Saale“ gefragt, dann dürften sich die Ergebnisse zumindest in ihrer Größenordnung nicht widersprechen. D.h. auch die Zahlungsbereitschaft derjenigen, die nur nach dem Gut „Erhalt der Flußauen an der Saale“ gefragt werden, dürfen keine signifikant größere Zahlungsbereitschaft nennen als diejenigen, die nach der Zahlungsbereitschaft für das Gut „Erhalt der Flußauen an der Saale“ als Bestandteil eines umfassenderen Schutzprogramms gefragt werden.

Anhand von Tabelle 8 lassen sich weiterhin zwei Begriffe verdeutlichen, die zur Differenzierung des Embedding-Effektes genutzt werden (vgl. HOEVENAGEL 1996: 59): Perfect- und Regular-Embedding. Von Perfect Embedding ist dann auszugehen, wenn die WTP für ein bestimmtes Gut ähnlich hoch ist wie die für ein umfassenderes Gut. Für die in der Tabelle dargestellten Programme würde dies bedeuten, daß $WTP_{K_1} = WTP_{K_2} = WTP_{K_3}$ für die auf der Diagonale angeordneten Güter ist. Die Zahlungsbereitschaft dürfte nur dann annähernd gleich sein, wenn die Güter auf den jeweiligen Ebene perfekte Substitute wären, d.h. der Einfluß auf ihre Wohlfahrt eines nationalen Programms zur Erhaltung von Flußauen derselbe wäre wie der eines Programms für die Erhaltung der Flußauen an der Elbe.

Regular Embedding liegt dann vor, wenn für ein Gut andere Zahlungsbereitschaften geäußert werden abhängig davon, ob die Zahlungsbereitschaft aus einem umfassenderen Gut abgeleitet wird oder das Gut allein für sich stehend bewertet wird. Die Werte für die „eingebetteten“ Güter (K_2 und K_3) sollten geringer sein als wenn die Güter unabhängig voneinander bewertet werden (z.B. nur K_3). Nach HOEVENAGEL hängt die Antwort darauf davon ab, wie die unabhängig bewerteten Güter definiert sind. Würde den Befragten aus Gruppe III gesagt, daß das Schutzprogramm an der Elbe das einzige ist,

das durchgeführt werden soll, dann wäre eine höhere Wertschätzung dieser Gruppe für K_3 konsistent gegenüber einer geringeren Wertschätzung in Gruppe I, die das Gut K_3 als Bestandteil eines umfassenderen Programmes präsentiert bekommt. Würde den Befragten aus Gruppe III aber das Programm K_3 als eines von mehreren Programmen zum Schutz von Flußauen dargestellt, dann müßte die Wertschätzung für alle K_3 horizontal in etwa gleich sein.

Tabelle 8: Part-Whole Bias beim Gut „Flußauen“

Umweltgut		Jeweilige Gruppe von Befragten		
		I	II	III
K_1	Erhalt der Flußauen in Deutschland	WTP_{K_1}		
K_2	Erhalt der Flußauen in den neuen Bundesländern	WTP_{K_2}	WTP_{K_2}	
K_3	Erhalt der Flußauen an der Elbe	WTP_{K_3}	WTP_{K_3}	WTP_{K_3}

Quelle: eigene Darstellung nach HOEVENAGEL (1996)

Die Diskussion um den Part-Whole Bias hat zu diversen Studien geführt, die sich insbesondere einer empirischen Überprüfung dieses Phänomens zugewandt haben. Einen guten Überblick über diese Studien gibt CARSON (1997). Er kommt zu der Einschätzung, daß die Hypothese des Part-Whole Bias (bei ihm *Scope Insensitivity*) in der überwiegenden Zahl der Studien zurückgewiesen werden konnte. In den Studien, in denen eine Zurückweisung dieser Hypothese nicht eindeutig möglich war, waren CARSON zufolge u. a. die Stichprobenumfänge zu klein oder es wurden Umfragen in Einkaufszentren oder per Telefon durchgeführt, so daß die Befragten sich dem Interview nicht mit der notwendigen Aufmerksamkeit widmen konnten. Insbesondere weist er daraufhin, daß in 19 der Studien auch Non-use Values ermittelt wurden und trotzdem die Hypothese des Part-Whole Bias zurückgewiesen werden konnte. Weitere empirische Belege dafür, daß der Part-Whole Bias bei entsprechender Ausgestaltung der Untersuchung nicht die verzerrende Wirkung auf die Ergebnisse hat, finden sich z. B. in ELSASSER (1996) und DEGENHARDT, GRONEMANN (1998). Letztere haben ihre Arbeit explizit der Erklärung des Embedding-Effektes gewidmet und bieten hierfür einige über die bisher in der Literatur diskutierten Erklärungsansätze an. Schließlich sei noch eine Untersuchung von BATEMAN ET AL. (1997) angeführt, in der untersucht wurde, ob der Part-Whole Bias auch bei privaten Gütern auftreten kann. Folgt man ihren Ergebnissen, dann muß davon ausgegangen werden, daß dieser Bias sehr wohl auch bei pri-

vaten Gütern auftritt und nicht nur als ein Spezialproblem der Bewertung öffentlicher Güter mit Hilfe der CVM angesehen werden kann.

In der Literatur werden weitere Arten möglicher Zuordnungsfehler unterschieden (vgl. ELSASSER 1996: 94f.), die hier kurz aufgeführt werden sollen:

- *Zeitbezug der Zahlung*: Vor allem bei periodischen Zahlungsverkehrsmitteln (Steuern, befristete Nutzungslizenzen) könnten Zahlungen auf abweichende oder nicht definierte Zeithorizonte bezogen werden. Beispiel hierfür wäre, daß sich die Befragten nicht ausreichend darüber im Klaren wären, daß es sich bei einer monatlichen Zahlungsbereitschaft um regelmäßig wiederkehrende Zahlungen handelt.
- *Geographische Fehlbezüge*: Sie können dann auftreten, wenn den Befragten der geographische Bezug bzw. die räumliche Abgrenzung des Gutes nicht deutlich ist (z. B. Elbauen nur ein Teil der gesamten Flußauen in Deutschland).
- *Personelle Fehlbezüge*: In diesem Fall würden Antworten auf andere Personen bzw. Gruppen bezogen als dies vom Fragesteller beabsichtigt war. Als typisch hierfür kann das Problem angesehen werden, zwischen individuellen Zahlungsbereitschaften und denen von Haushalten zu unterscheiden.
- *Kategorien von Gütern bzw. Güterbündeln und Nutzen*: Dieser Zuordnungsfehler könnte dann vorliegen, wenn die Befragung z. B. auf den Freizeitnutzen abzielen würde, aber die Befragten auch Nutzenquellen wie Existenzwert oder ökologische Leistungen einbeziehen.
- *Politischer Kontext*: Dieser Fehler kann schließlich dann auftreten, wenn eine spezifische Maßnahme als Bestandteil eines Politikprogramms aufgefaßt wird und folglich das gesamte Programm und nicht die spezifische Maßnahme bewertet wird (Beispiel: Luftreinhaltepolitik als Bestandteil der gesamten Umweltpolitik).

5.7.6 **Beteiligungsproblem (*Nonresponse Bias*) und Behandlung von „Ausreißern“**

Insbesondere bei schriftlichen Befragungen ist es häufig ein Problem, daß nicht alle Befragten antworten. Eine Erklärung hierfür könnte darin liegen, daß die Antwortenden ein höheres Interesse am Gegenstand der Befragung besitzen als diejenigen, die nicht antworten. Dadurch entstehen aber Probleme mit der Repräsentativität der Stichprobe. Damit ist dann das Problem verbunden, daß es zu einer Überschätzung des ökonomischen Wertes kommen kann, wenn insbesondere Haushalte oder Personen mit höherem Bildungsniveau und/oder höherem Einkommen sich an der Umfrage beteiligen und so die durchschnittliche Zahlungsbereitschaft höher liegt als sie eigentlich bei einer repräsentativen Beantwortung gewesen wäre.

Gleichzeitig stellt sich aber die Frage, ob die tatsächliche Wertschätzung für ein Gut dadurch nicht auch unterschätzt werden kann. Es könnte sein, daß die Verweigerung einer Antwort bzw. die Nicht-Beteiligung an einer schriftlichen Befragung auf Protestverhalten zurückzuführen ist. In diesem Fall hängt es von den Protestgründen ab, ob davon ausgegangen werden kann, ob tatsächlich keine Wertschätzung für das Gut besteht, eine unendlich hohe Wertschätzung (lexikographische Präferenzen: vgl. SPASH, HANLEY 1995 sowie das Kapitel 5.7.7) oder aber von einer „normalen“ bzw. durchschnittlichen Zahlungsbereitschaft ausgegangen werden kann. So ist der Fall denkbar, daß eine Antwort mit der Begründung verweigert wird, daß die Bereitstellung des jeweiligen Gutes Aufgabe des Staates sei. In diesem Fall kann durchaus eine Wertschätzung für das zu bewertende Gut bestehen, allerdings wird diese durch die Untersuchung nicht erfaßt. Um eine Zuordnung der Protestantworten zu ermöglichen, werden in etlichen Studien Filterfragen eingebaut. Damit wird versucht, zu ermitteln, warum die Befragten nicht bereit waren, eine Zahlungsbereitschaft zu nennen (vgl. HOEVENAGEL 1996: 67). So wird z. B. danach gefragt, ob generell kein Interesse an dem zu bewertenden Umweltgut besteht, die derzeitige Einkommenssituation keine Zahlung ermöglicht oder ob kein Vertrauen gegenüber dem Zahlungsempfänger besteht. Je nachdem, wie konservativ die Abschätzung sein soll, können zum einen Personen, die eine Antwort auf die Frage nach der Zahlungsbereitschaft oder die Teilnahme an der Befragung generell abgelehnt haben, aus dem Sample herausgenommen werden. Zum anderen kann für diejenigen, die keine Zahlungsbereitschaft geäußert haben, für die aber aufgrund ihrer Antwort auf die Filterfragen eine positive Wertschätzung angenommen werden kann, die durchschnittliche Zahlungsbereitschaft „unterstellt“ werden.

Darüber hinaus stellt sich das Problem, wie mit Ausreißern umgegangen werden soll. Hierzu gehören insbesondere Wertschätzungen, die deutlich über der tatsächli-

chen Zahlungsbereitschaft - und zum Teil auch der Zahlungsfähigkeit - liegen dürften. Zur Aufdeckung vermeintlicher Ausreißer bietet sich ein Abgleich mit dem von der Person genannten Einkommen an. Macht die geäußerte Zahlungsbereitschaft einen Großteil des Einkommens aus, oder liegt gar darüber, dann ist davon auszugehen, daß es sich um Protestantworten oder strategische Antworten handelt. In einigen Studien wurden Beträge definiert, oberhalb derer die genannten Zahlungen nicht mehr als „richtiger“ Ausdruck gewertet werden. So hat HOEVENAGEL (1996: 67) in einer von ihm durchgeführten Untersuchung festgelegt, daß genannte Zahlungsbereitschaften oberhalb von zehn Prozent des jeweiligen Einkommens abgeschnitten werden.

5.7.7 Problematische Annahmen über die Präferenzen

Über die beiden in Abschnitt 4.1 genannten Voraussetzungen für den Umgang mit Präferenzen hinaus werden in der neoklassischen Ökonomie weitere Annahmen getroffen, die insbesondere für die Bewertung mit Hilfe der Contingent Valuation Method von Bedeutung sind: Dies ist zum einen die Annahme, daß die Präferenzordnung vollständig und transitiv ist. Zum anderen ist dies die Annahme, daß die Präferenzordnung stetig ist. Für den Fall einer unvollständigen und nicht transitiven Präferenzordnung kann es dazu kommen, daß Präferenzen „konstruiert“ werden. Ist die Präferenzordnung nicht stetig, dann können lexikographische Präferenzen vorliegen. Wenn davon ausgegangen werden muß, daß die grundlegenden Annahmen der neoklassischen Ökonomie nicht die Struktur der bei den Individuen tatsächlich vorhandenen Präferenzen beschreiben, dann ist fraglich, inwieweit die mit Hilfe der Monetarisierung ermittelten Werte in Kosten-Nutzen-Analysen eingestellt werden können. Auf beide Fälle soll hier kurz eingegangen werden.

Konstruierte Präferenzen

Von der ökonomischen Theorie wird angenommen, daß die Individuen über eine vollständige Präferenzordnung verfügen. Daran anknüpfend geht die Ökonomie von der Annahme gegebener Präferenzen aus, d.h. auch für Güter wie z. B. Umweltgüter, die noch nicht auf Märkten gehandelt werden, kennt das Individuum schon seine Wertschätzung und kann diese in monetären Größen angeben. „In other words, it is assumed that people have true hidden preferences for environmental changes and that they are capable of transforming them into monetary units“ (HOEVENAGEL 1992 177). Aufgabe der Ökonomie wäre es dann nur noch, diese gegebenen, aber bisher versteckten Präferenzen durch den Einsatz geeigneter Techniken „ans Tageslicht“ zu bringen.

Doch zeigen die Ergebnisse entsprechender Experimente deutlich, daß von gegebenen, „nur“ versteckten Präferenzen nicht generell ausgegangen werden kann. Wären

sie gegeben, dann müßten z. B. die in entsprechenden Situationen geäußerten Präferenzen für bestimmte Umweltgüter unabhängig von den jeweiligen Umständen der Präferenzermittlung gleich sein. Tatsächlich hängt die geäußerte Zahlungsbereitschaft aber stark von den Umständen der Präferenzermittlung ab. Zurückgeführt wird dies darauf, daß die Präferenzordnung der Individuen nicht vollständig, sondern vielmehr unvollständig ist und Präferenzen für das zu bewertende Gut überhaupt erst einmal gebildet – konstruiert - werden müssen. „Die Theorie kümmert sich nicht um die Frage, wie Präferenzen entstehen, aber sie setzt voraus, daß sie entstanden sind. Ist dies eine sinnvolle Annahme? Es sei an dieser Stelle die Vermutung erlaubt, daß die Annahme gegebener Präferenzen nur in Kontexten Sinn macht, in denen Märkte für private Güter existieren. Im Falle öffentlicher Güter oder externer Effekte verliert sie ihre Überzeugungskraft, und zwar aus dem folgenden Grund: Es ist eine Alltagserfahrung, die vielfach experimentell bestätigt worden ist, daß Menschen nicht a priori über Präferenzen (erst recht keine konsistente Präferenzordnung) verfügen. Präferenzordnungen werden ad hoc gebildet, wenn sie benötigt werden“ (WEIMANN 1996a: 438).

Somit ist eine Unterscheidung zwischen privaten und öffentlichen Güter in bezug auf die Beurteilung der Präferenzordnung wesentlich: Während im Fall privater Güter viel eher davon ausgegangen werden kann, daß gegebene Präferenzen vorliegen, dürfte dies bei öffentlichen Gütern weit weniger der Fall sein. Allerdings weisen ARROW (1993) am Beispiel des Kaufs eines Autos und WEIMANN (1996a) am Beispiel des Kaufs eines Hauses darauf hin, daß bei komplexen privaten Gütern, die zudem nur selten erworben werden – zumindest im Vergleich zu eher alltäglichen erworbenen Gütern – ähnliche Phänomene auftreten können. Für die Bewertung öffentlicher Güter ergibt sich daraus die Notwendigkeit, Informationen über die Präferenzen bzw. diejenigen Größen zu gewinnen, die Einfluß auf die geäußerte Zahlungsbereitschaft haben. Denn wenn angenommen wird, daß Präferenzen konstruiert werden, dann ist es zentral zu wissen, „woraus“ sie konstruiert wurden und ob die geäußerten Zahlungsbereitschaften mit dem von den Ökonomen unterstellten Verständnis von Präferenzen vereinbar sind.²⁰

Als ein Ansatz, der in der Forschung in diesem Zusammenhang zunehmend an Bedeutung gewinnt, ist die Messung sog. Einstellungen (*Attitudes*) im Rahmen der Contingent Valuation Method zu nennen. Der Begriff der Einstellungen hat in der (Sozial-) Psychologie im Laufe der Zeit mit dem Auf und Ab verschiedener Schulen und Erklä-

²⁰ Unter anderem angestoßen durch die Diskussion um die Güte der Ergebnisse von CVM-Studien ist eine Vielzahl von Beiträgen zur psychologischen Erklärung der geäußerten Zahlungsbereitschaften erschienen, die zum Teil an die ältere Diskussion um die Rationalitätsannahme der Ökonomie anknüpfen (vgl. z. B. KAHNEMAN ET AL. (1993), SLOVIC (1995), für einen breiteren Überblick über das Verhältnis von Psychologie und Ökonomie siehe den jüngst erschienenen Überblicksartikel von RABIN (1998)).

rungsansätze unterschiedliche Deutungen und eine Vielzahl von Definitionen erfahren.²¹ Nach STADE sind soziale Einstellungen „erlernte, relative überdauernde Wahrnehmungsorientierungen und Reaktions- bzw. Handlungsbereitschaften; sie sind verhaltenswirksam und dabei explizit evaluativ, d.h. bewertend auf eine Klasse sozialer Objekte bezogen“ (zitiert nach SPADA 1990). In den Befragungen wird daher nicht nur nach der Zahlungsbereitschaft gefragt, sondern es wird auch versucht, Einstellungen der Befragten gegenüber dem Bewertungsobjekt zu messen. Hierfür werden ihnen Aussagen²² präsentiert, die sie mit Hilfe einer vorgegebenen Bewertungsskala (z. B. Likert-Skala, semantisches Differential) bewerten sollen. Die hierdurch gewonnenen Indikatoren werden dazu genutzt, zu prüfen, ob die geäußerten Zahlungsbereitschaften eher von ökonomischen Größen (z. B. Einkommen) oder aber eher von eben den Einstellungen der Befragten bestimmt werden. Die Ergebnisse bisheriger Arbeiten sind dabei durchaus unterschiedlich: Während AJZEN ET AL. (1992) zu dem Ergebnis kommen, daß in ihrer Untersuchung die Einstellungen die geäußerte Zahlungsbereitschaft dominierten, kam WIERSTRA (1997) zu dem Ergebnis, daß es sehr stark von Typus des jeweiligen Gutes abhängt, ob eher ökonomische Größen oder Einstellungen die geäußerte Zahlungsbereitschaft dominieren (vgl. hierzu z. B. die Arbeiten von AJZEN und DRIVER (1992), MCCELLAND (1997), WIERSTRA (1997)). Die Messung von Einstellungen kann als ein vielversprechender Ansatz angesehen werden, um einer Beantwortung der von WEIMANN (1996a: 438) gestellten Frage näher zu kommen, was eigentlich gemessen wird, wenn nicht von gegebenen Präferenzen ausgegangen werden kann.

Lexikographische Präferenzen

Eine Präferenzordnung wird als lexikographisch bezeichnet, wenn ein Individuum die Güterbündel zunächst ausschließlich nach der Menge eines Gutes beurteilt und erst danach die Mengen der anderen Güter von Bedeutung sind (vgl. FEES-DÖRR 1991: 191). Der Name lexikographische Präferenzen leitet sich aus einer Analogie zum Lexikon ab, in dem alle Wörter mit dem Anfangsbuchstaben A vor allen Wörtern mit dem Anfangsbuchstaben B kommen, egal welcher Buchstabe an zweiter Stelle folgt (Azz vor Baa).

Liegen lexikographische Präferenzen vor, dann wäre die Stetigkeitsannahme verletzt. In diesem Fall würden gar keine physisch unterschiedlichen Güterbündel vorlie-

²¹ Für einen kurzen Überblick vgl. GÜTLER (1996).

²² Beispiele für derartige Aussagen sind: „Mich beunruhigt, daß viele Tier- und Pflanzenarten in Deutschland vom Aussterben bedroht sind“, „Der Verlust an Natur und Landschaft läßt sich durch einen Zuwachs bei anderen Gütern wieder ausgleichen“ oder „Es ist fair, wenn jeder einen angemessenen Beitrag für den Schutz von Natur und Landschaft zahlt“.

gen, die gleich beurteilt werden. Entweder gleichen sich die Güterbündel vollständig, oder es wird unabhängig von den Mengen aller nachgeordneten Güter das Güterbündel mit der größeren Menge des Gutes vorgezogen, das in der lexikographischen Ordnung am höchsten rangiert. Anders ausgedrückt: Die Individuen sind nicht bereit, ein bestimmtes Gut durch ein anderes Gut zu substituieren bzw. zu ersetzen, unabhängig von der Menge, die ihnen für die Aufgabe des einen Gutes geboten wird. Übertragen auf den Bereich der monetären Bewertung von Natur und Landschaft heißt dies, daß Eingriffe in die Umwelt kategorisch abgelehnt werden. Als Beispiel hierzu wird z. B. angeführt, daß eine Verringerung im Bestand einer Tier- und Pflanzenart als Folge eines Eingriffes hingenommen wird, wenn dieser Verlust entsprechend kompensiert wird. Die Kompensation kann z. B. durch eine Steigerung des Konsums von Marktgütern erfolgen. Für den Fall aber, daß die Tier- und Pflanzenart durch den Eingriff ausgerottet würde, wäre der Verlust aus Sicht der bewertenden Individuen nicht mehr kompensierbar. In diesem Fall kann keine noch so hohe Steigerung des Konsums von Marktgütern den Verlust kompensieren.

In einigen wenigen Untersuchungen wurde bisher die Existenz lexikographischer Präferenzen geprüft. Zu nennen sind hier STEVENS ET AL. (1991) und SPASH und HANLEY (1995). Während STEVENS et al. es zum Ziel hatten, vier verschiedene Tierarten Neuenglands zu bewerten, haben SPASH und HANLEY versucht, den Wert von Restbeständen des Caledonischen Kiefernwaldes in Schottland zu ermitteln. Diese immer schneller verschwindenden Wälder stellen das Habitat für einige seltene Vogelarten sowie Säugetiere dar. Beide Studien kommen zu dem Ergebnis, daß etwa ein Viertel der Befragten lexikographische Präferenzen gezeigt hätte, daß heißt einen Trade-off zwischen einem Geldbetrag und der Frage, ob die gefährdeten Arten erhalten bleiben sollen, abgelehnt haben.

Die Frage ist nun, welche Konsequenzen lexikographische Präferenzen für die ökonomische Bewertung haben. Stellt sie nach Meinung einiger Autoren eine Grenze für die ökonomische Bewertung dar (vgl. KOSZ 1997b), so halten MARGGRAF und STREB dem entgegen, daß dies nicht richtig sei. Nach ihrer Ansicht ist die Bereitschaft zum Tausch zwischen zwei Gütern keine Voraussetzung des ökonomischen Wertbegriffs. „Der Kern des ökonomischen Wertbegriffs umfaßt die Vorstellung, den Wert einer Umweltveränderung als minimale Kompensationsforderung oder maximale Zahlungsbereitschaft zu bestimmen und beinhaltet keine Aussage darüber, wie groß diese Beträge sein müssen“ (MARGGRAF, STREB 1997: 227).

Für eine Studie zur ökonomischen Bewertung insbesondere von Natur und Landschaft, wie dies auch im Fall der Stromlandschaft Elbe der Fall wäre, stellt sich daher die Aufgabe, die Möglichkeit dieser Form von Präferenzstruktur im Untersuchungsdesign zu berücksichtigen. Da bei Vorliegen lexikographischer Präferenzen ein Tausch

abgelehnt wird, bietet allein die CVM die Möglichkeit, diese Präferenzen aufzudecken. Durch entsprechende Fragen könnte aufgedeckt werden, ob von den Individuen die Zahlung für eine Verbesserung oder sogar eine Kompensation für eine Verschlechterung der Umweltqualität abgelehnt wird. In diesem Fall wäre zu klären, wie mit den Werten im Rahmen einer Kosten-Nutzen-Analyse umzugehen ist und welche Empfehlungen sich daraus an die Politik ergeben. Als ungeklärt muß es bis dato aber ebenfalls angesehen werden, ob die lexikographischen Präferenzen nicht zumindest zum Teil auch der hypothetischen Situation der CVM zuzuschreiben sind: „Individuals who claim that no amount of money would compensate them for any environmental change, no matter how small, in a hypothetical market might well change their position if offered real cash to accept a small degradation in environmental quality“ (HANLEY ET AL. 1995: 261).

5.7.8 Die Richtlinien des NOAA-Panels

Das NOAA-Panel kam in seinem Schlußbericht zu dem Ergebnis, daß mit Hilfe der Contingent Valuation Method sehr wohl auch die sog. Non-use Values für die Feststellung von Schäden an Umweltgütern ermittelt werden können (ARROW ET AL. 1993). Allerdings seien bei der Durchführung der CVM bestimmte Richtlinien zu berücksichtigen, um die Validität der Ergebnisse zu gewährleisten. Das Panel hat einen Katalog von Richtlinien vorgestellt, dessen Berücksichtigung die Zuverlässigkeit weitgehend „gewährleisten“ soll. In WILLIS (1995) findet sich u.a. ein Überblick darüber, welche dieser Richtlinien zur Zeit in welchem Umfang Anwendung in welchen Contingent Valuation Studien finden. Hier soll eine Auswahl dieser Richtlinien aufgeführt werden, um dann kurz auf die durch sie ausgelöste Diskussion einzugehen.

NOAA - Richtlinien für die Durchführung von CVM-Studien

1. Generelle Richtlinien

- *Stichprobenart und -größe*: Die Verwendung von Zufallsstichproben ist notwendig.
- *Antwortverweigerungen*: Sie sollten minimiert werden, da Ergebnisse sonst unzuverlässig werden.
- *Persönliche Befragung*: Es ist unwahrscheinlich, daß verlässliche Schätzungen der Werte durch schriftliche Befragungen ermittelt werden können. Persönliche Interviews sind vorzuziehen, obwohl telefonische Interviews Vorteile bezüglich der Kosten und einer zentralen Überwachung bieten.
- *Pretest auf Interviewereffekt*: Die Interviewer können zu einer Verzerrung in Richtung auf das „sozial erwünschte“ Verhalten führen, da die Erhaltung der Natur vielfach als etwas Positives angesehen wird. Contingent Valuation (CV) Studien sollten Experimente enthalten, die solche Interviewereffekte abschätzen.

- *Dokumentation*: Jede CV Studie sollte dokumentieren, welche Stichprobe erhoben wurde, das Verfahren zur Stichprobenziehung, die Größe der Stichprobe und die Rate der Antwortverweigerung. Der verwendete Fragebogen sollte präsentiert werden und die Daten anderen Forschern zur Verfügung gestellt werden
- *Pretest des Fragebogens*: Durch entsprechende Pretests soll sichergestellt werden, daß die Antwortenden die Beschreibungen und Fragen ausreichend verstanden haben.

2. Richtlinien für Studien zur Wertermittlung

- *Konservatives Design*: Dies erhöht die Zuverlässigkeit, da extreme Antworten, die die geschätzten Werte erhöht, wegfallen. Es wird die Alternative bevorzugt, die die Zahlungsbereitschaft unterschätzt.
- *Frageformat*: Es sollte die Zahlungsbereitschaft (Willingness to pay) verwendet werden statt der Verkaufsbereitschaft (Willingness to sell), da WTP die konservative Alternative ist.
- *Entscheidungsformat*: Die Bewertungsfrage sollte wie eine Abstimmungsfrage gestellt werden.
- *Akkurate Beschreibung von Programm oder Politik*: Der Befragte muß ausreichende Informationen über das zu bewertende Umweltprogramm erhalten.
- Erinnerung an *ungeschädigte Substitute*
- *Angemessener Zeitabstand zu eingetretenen Unfällen*: Dadurch sollen Fehleinschätzungen bezüglich der Möglichkeiten der Wiederherstellung vermieden werden, da Befragte oft einen substantiellen Verlust an passivem Nutzen angeben, auch wenn sie darüber informiert werden, daß eine vollständige Wiederherstellung später eintreten wird.
- *Ja/Nein Folgefragen*: Um zu erfassen, warum Befragte mit ja oder nein auf die Frage nach der Zahlungsbereitschaft geantwortet haben

3. Ziele für Studien zur Wertermittlung

- *Alternative Ausgabemöglichkeiten*: Befragte sollten daran erinnert werden, daß ihre Zahlungsbereitschaft für Umweltgüter ihre Konsummöglichkeiten für andere Güter reduziert.
- *Verhinderung des Warm Glow Effektes*: Die Untersuchung sollte so gestaltet sein, daß nicht der Nutzen aus dem Akt des Gebens gemessen wird, sondern der Nutzen aus der Veränderung des Güterangebotes.
- *Dauerhafte oder vorläufige Verluste*: Befragte sollten zwischen diesen beiden unterscheiden können, da vollständige Wiederherstellung in der Zukunft passive Nutzungsverluste erheblich reduziert.
- *Kalkulationen des Gegenwartswertes bei vorläufigen Verlusten*: Es sollte dargestellt werden, daß Befragte bezüglich auf den Wiederherstellungszeitpunkt deutlich reagieren.
- *Nachweis der Zuverlässigkeit*: Derjenige, der CVM-Studie entwirft, hat die Aufgabe, die Zuverlässigkeit seiner Ergebnisse plausibel zu machen.

Quelle: eigene Übersetzung nach WILLIS (1995)

Die ENVIRONMENTAL PROTECTION AGENCY (EPA) hat neben Anderen relativ deutliche Kritik an den vom NOAA-Panel aufgestellten Kriterien geübt (vgl. EPA 1994). Hauptmotiv dafür war, daß durch die Regelungen die Kosten für eine CVM deutlich in die Höhe getrieben werden. Zum einen sei damit die Gefahr verbunden, daß als Folge nur noch eingeschränkt von diesem Instrument Gebrauch gemacht wird. Dies hätte aber zur Konsequenz, daß die Non-use Values in vielen Entscheidungsprozessen nicht berücksichtigt würden und damit die Schäden an den natürlichen Ressourcen unterschätzt würden. Zum anderen sei es aber auch fraglich, wie stark durch die Regeln des NOAA-Panels tatsächlich die Qualität der Untersuchungsergebnisse verbessert würde. So gesehen müßten die Regeln selbst einer Kosten-Nutzen-Analyse unterzogen werden. Schließlich wird von seiten der EPA (1994) befürchtet, daß die weitere Entwicklung der Methode zum Stillstand kommt, da die Regeln als Standard angesehen werden und Untersuchungen, die sich nicht daran orientieren, nur geringe Akzeptanz bei den Entscheidungsträgern finden (vgl. hierzu auch SCHULZE 1996). Es stellt sich somit die Frage, ob aus Sicht der methodischen Entwicklung schon der Zeitpunkt gegeben ist, Verfahrensregeln festzuschreiben und damit zumindest die Wahrscheinlichkeit methodischer Verbesserungen zu verringern.

Insbesondere die Frage nach dem Verfahren zur Präferenzenthüllung hat eine kontroverse Diskussion ausgelöst. Das NOAA-Panel hatte das dichotome Frageformat als dasjenige festgelegt, das für die Ermittlung „gesicherter“ Werte für die Wertschätzung vorzuziehen sei. Allerdings sprechen auch etliche Gründe für den Einsatz anderer Formate, so daß aus der wissenschaftlichen Diskussion keine eindeutige Empfehlung abzuleiten ist. So dürfte die offene Frage - zusammen mit Zahlungskarten - insbesondere aufgrund ihres geringeren Aufwandes und dadurch auch geringeren Kosten häufig Anwendung finden²³. Darüber hinaus ist zu berücksichtigen, daß länderspezifische Gegebenheiten eine Rolle spielen. So weist WILLIS (1995: 127) darauf hin, daß „...people in Britain are unfamiliar with voting on tax propositions, compared with people in the USA where they are a regular feature of state election process“. Weiterhin ist davon auszugehen, daß auch die dichotome Frageform zu Verzerrungen führen kann. Liegt z. B. der dem Befragten genannte Betrag nur knapp über dem, den er eigentlich zu zahlen bereit ist, dann kann er geneigt sein, in diesem Fall zuzustimmen. Einige Studien weisen daraufhin, daß die dichotome Frageform zu signifikant höheren Zah-

²³ Nach Kenntnis des Autors haben bis auf wenige Ausnahmen alle bisher im deutschsprachigen Raum durchgeführten und veröffentlichten Studien - ohne allerdings Anspruch auf Vollständigkeit zu erheben - mit offenen Fragen bzw. mit Zahlkarten gearbeitet: so u.a. HAMPICKE et al. (1991), PRUCKNER (1994), LÖWENSTEIN (1994), ZIMMER (1994), ELSASSER (1996), KÄMMERER et al. (1996), SCHÖNBÄCK et al. (1997). Ausnahmen bilden JUNG (1995) und HACKL (1997).

lungsbereitschaft führt (vgl. u.a. die Ausführungen bei BATEMAN et al. 1993, GEISENDORF et al. 1996).

Letztlich gibt es für die Auswahl der Frageform keine grundsätzlichen Vorgaben, wie WILLIS (1995: 127) schreibt. Nach ihm spricht vor allem die Verwendung großer Stichproben sowie der höhere statistische Aufwand häufig gegen den Einsatz der dichotomen Frageform. BATEMAN ET AL. (1993: 41) ziehen zwar aus ihrer Untersuchung den Schluß, daß die Befragten große Unsicherheit bei der Beantwortung der open ended-Frage gezeigt hätten, doch kann dieses Ergebnis nicht als dominante Begleiterscheinung der open ended-Frage angesehen werden. Zumindest deuten die Ergebnisse vieler Studien nicht darauf hin bzw. geben die jeweiligen Autoren dies nicht als ein vorrangliches Problem ihrer Untersuchung an. Im Gegenteil: In der Literatur erlebt die open ended-Fragestellung zur Zeit sogar wieder eine gewisse Renaissance (vgl. hierzu HANLEY ET AL. 1996, READY ET AL. 1996). Somit geben die vom NOAA-Panel aufgestellten Richtlinien auf der einen Seite eine Orientierung und stellen sicher auch einen Maßstab für die Beurteilung empirischer Studien dar. Doch können die Richtlinien derzeit in ihrer Gesamtheit keinesfalls als allgemeingültige Standards angesehen werden. Vielmehr ist bei der Durchführung von CVM-Studien jeweils zu prüfen, ob und welche Richtlinien angewendet bzw. nicht angewendet werden.