

## 2 Theoretischer Hintergrund

### 2.1 Besonderheiten der Umwelt als Gut

Die traditionelle Vorgehensweise in der Ökonomie, den Nutzen von Gütern und Services zu bestimmen, besteht in dem Heranziehen von Marktinformationen zur Ableitung von Preis-Absatz-Funktionen. Die Preis-Absatz-Funktion zeigt auf, wie sich die Nachfrage nach einem Gut in Abhängigkeit vom Preis verändert (Homburg & Krohmer, 2006). Im Fall, dass für ein Gut (noch) kein Markt existiert, kann der Nutzenwert jedoch nicht vom Marktverhalten erschlossen werden. Ein Markt existiert z. B. nicht für neue oder in der Planung begriffene Konsumgüter sowie für kollektive Güter wie die Umwelt.

Kollektive Umweltgüter wie Gewässer- und Luftqualität sind im Vergleich zu Konsumgütern durch eine Nicht-Rivalität im Konsum und durch eine Nicht-Ausschliessbarkeit der Nutzung gekennzeichnet (Olson, 2009). Keine Person kann vom Konsum einer verbesserten Luftqualität ausgeschlossen werden und die Menge der konsumierten Luft von einer Person reduziert nicht die Menge, die einer anderen Person zur Verfügung steht. Weitere Beispiele für kollektive Umweltgüter sind saubere Böden, intakte Wälder, unbebaute Naturlandschaften. Bei privaten Gütern wie Autos oder Nahrungsmittel sind die Kriterien der Nicht-Rivalität und der Nicht-Ausschließbarkeit hingegen verletzt. Bei einem Brot z.B. führt der Konsum einer von zwei Personen dazu, dass die zweite Person weniger Brot hat (Rivalität). Ein Auto ist nur für die Personen zugänglich, die den Preis des Herstellers bezahlen (Ausschließbarkeit). Aus diesem wesentliche Unterschied zwischen kollektiven und privaten Gütern ergeben sich problematische Ungleichgewichte hinsichtlich des Konsums und der Bereitstellung von Umweltgütern (Esser, 2002).

Da der Nutzen von einem Umweltgut allen zugutekommt und nicht auf bestimmte Konsumenten zu beschränken ist, haben private Anbieter keinen Anreiz das Gut anzubieten; es entsteht ein Bereitstellungsproblem (Ostrom, 1990). Während der Markt bei privaten Gütern durch den Preismechanismus eine Steuerungsfunktion hat, die eine effiziente Bereitstellung gewährleistet, kommt es bei vielen gesellschaftlichen Gütern zu einem s.g. Marktversagen (vgl. Musgrave, Musgrave & Kullmer, 1994). Mit anderen Worten, der Konsum und die Bereitstellung von Umweltgütern geraten aus dem Gleichgewicht.

Marktversagen entsteht dann, wenn die Märkte die tatsächlichen Kosten für das Gut bzw. den tatsächlichen Nutzen vom Gut nicht widerspiegeln. Ein Beispiel dafür wäre der Benzinpreis, der nicht ganz für die Kosten aufkommt, die der Gesellschaft durch die Abgase bei der Verbrennung entstehen. Ein weiteres Beispiel wäre eine Pflanzenart in den Amazonas mit

einem hohen potentiellen Wert für die Medizin, die infolge von ausgelassenen Schutzmaßnahmen stark bedroht oder ausgestorben ist. Das Marktversagen geht mit einem übermäßigen Konsum (Luftverschmutzung und Klimaerwärmung durch die Autoabgase) bzw. mit einer Unterproduktion knapper gesellschaftlicher Ressourcen (unzureichender Amazonas Schutz) einher. Beispiele für die Folgen des Marktversagens im Umweltbereich sind die Zerstörung der Ozonschicht, die Überfischung der Meere, die Abholzung der Urwälder, die globale Klimaerwärmung und das Aussterben vieler Arten.

Da der Markt keine effiziente Bereitstellung von Umweltgütern gewährleistet, werden üblicherweise Entscheidungen über ihre Bereitstellung bzw. über das Ergreifen von Schutzmaßnahmen auf verschiedenen politischen Ebenen getroffen und realisiert. Dabei stellt sich oft die Frage, die Bereitstellung von welchen kollektiven (Umwelt-) Gütern Priorität haben soll (Umweltschutz oder Innenverteidigung; der Schutz einer bedrohten Tierart oder die Renaturisierung einer Landschaft). Bei alternativen Umweltschutzmaßnahmen wie bspw. im Falle verschiedener Renaturisierungsprogramme soll beurteilt werden, welche Alternative die effizienteste ist und daher realisiert werden soll. Beim Abwägen von ökonomischen Entwicklungsprogrammen ist ebenfalls zu entscheiden, welches Gut einen höheren Nutzen hat – die biologische Artenvielfalt in einem Wald oder die Abholzung der Bäume und die landwirtschaftliche Nutzung der Fläche. Als Grundlage für solche Entscheidungen ist erforderlich, den ökonomischen Wert von Umweltgütern zu kennen bzw. diesen zuverlässig zu bestimmen.

Der Wert eines (Umwelt-) Gutes kann in Anlehnung an die subjektive Werttheorie (vgl. Hens & Pamini, 2008) über den Nutzen erfasst werden, den das Gut stiftet. Der Nutzen setzt sich aus zwei Komponenten zusammen. Der direkte oder der Gebrauchsnutzen ergibt sich aus der direkten Inanspruchnahme des Umweltgutes wie z. B. in einem See baden. Der passive oder Nicht-Gebrauchsnutzen bezieht sich auf den Wert des Gutes, ohne dass dieses genutzt wird. Zu der letzteren Komponente zählen der Existenz-, der Options- und der Vererbungsnutzen (Freeman III, 2003). Der Existenzwert ergibt sich aus dem Wissen heraus, dass eine Umweltressource existiert bzw. diese geschützt wird ohne eine Nutzungsabsicht (Krutilla, 1967). Der Optionswert hingegen spiegelt den Wert für eine zukünftige Nutzungsmöglichkeit wider. Der Vererbungsnutzen bezeichnet die Wertschätzung dafür, dass zukünftige Generationen das Gut nutzen könnten.

Für die Existenz von Nicht-Gebrauchswerten im Umweltbereich spricht z. B. das Verhalten von Personen, die Beiträge in Naturschutzorganisationen oder für Naturschutzprogramme zwecks Erhaltung von seltenen Tierarten spenden, auch wenn sie keinen direkten Nutzen davon haben. Durch ihre Beiträge signalisieren sie jedoch einen positiven passiven Nutzen, der ihnen der Erhalt der seltenen Tierart stiftet. Der ökonomische Gesamtwert eines Umweltgutes ergibt sich somit aus der Summe der direkten und der passiven Nutzenwerte (Bateman et al., 2002).

Die meisten Entscheidungstheorien nehmen an, dass der Nutzen von verschiedenen Optionen die Präferenzen für diese bzw. die Wahlen zwischen diesen bestimmen (Keeney & Raiffa, 1976; Luce & Suppes, 1965; Luce & Tuckey, 1964). In den vergangenen 50 Jahren wurde eine Reihe von Methoden entwickelt, die auf der Basis von Präferenzen der Bevölkerung die Schätzung des ökonomischen Nutzens vom Schutz und von der Verbesserung der Umwelt ermöglichen. Die in der ökonomischen Umweltbewertung eingesetzten Methoden zur Nutzenmessung unterscheiden sich darin, in welchem Ausmaß beide Nutzenkomponenten eines Umweltgutes durch diese erfasst werden (vgl. Abschnitt 2.2). Gemeinsam an diesen Methoden ist, dass sie den Nutzen monetär in Form von Zahlungsbereitschaften erfassen. Als maximale Zahlungsbereitschaft bzw. Willingness To Pay wird der Geldbetrag aufgefasst, bei dem das Nutzenniveau der Personen vor dem Eintritt einer Veränderung in der Menge/Qualität des Umweltgutes und nach dem Eintritt der Veränderung unverändert bleibt (Alberini & Kahn, 2006). In der Sprache der Mikroökonomie handelt es sich dabei um die s.g. Hicksischen Variationsmaße der kompensierenden und der äquivalenten Variation (Mas-Colell et al., 1995).

## 2.2 Klassifikation der Methoden zur Messung von Zahlungsbereitschaften

Die einzelnen Methoden zur Messung von Zahlungsbereitschaften lassen sich jeweils in der Gruppe der *revealed* und *stated* Präferenzmethoden zusammenfassen (Bateman et al., 2002). Gemeinsam an diesen Methoden ist: a) dass diese auf Präferenzen für verschiedene Güter basieren bzw. auf Präferenzen für unterschiedliche Qualitäten/Quantitäten des Gutes; und b) der Nutzen in monetären Einheiten gemessen wird. Methodisch besteht der Hauptunterschied zwischen den *revealed* und den *stated* Präferenzmethoden darin, ob reales Verhalten oder Verhalten auf hypothetischen Märkten als Grundlage für die Schätzungen des Nutzens dient. Die Gruppe der *revealed* Präferenzmethoden enthält Techniken, die den Nutzen aus realem (Kauf-) Verhalten ermitteln. Im Marketingbereich zählen dazu die Analysen von Marktdaten, Labor- und Feldexperimenten sowie verschiedene Auktionen und Lotterien (vgl. Breidert, Hahsler & Reutterer, 2006; Voelckner, 2006 für eine Übersicht). Kennzeichnend für diese Methoden ist, dass die Personen für den Erwerb bzw. die Nutzung von einem Gut tatsächlich bezahlen müssen. Die Vickrey-Auktion (Vickrey, 1961) und die BDM-Lotterie (Becker, DeGroot & Marschak, 1964) finden ebenfalls in der Umweltbewertungsforschung z. B. im Rahmen von Validitätsstudien Anwendung (s. Abschnitt 2.5.2).

Im Bereich der Umweltbewertung gehören zu den populärsten *revealed* Präferenzmethoden die hedonische Preismethode und der Reisekostenansatz (Braden & Kolstad, 1991). Diese basieren auf beobachtbarem Verhalten zu einem Konsumgut, das in enger Beziehung zu einem Aspekt des Umweltgutes steht (Freeman III, 2003). Beispiele dafür wären die Reise zu einem Nationalpark und die damit verbundenen Kosten, die Wahl einer „grünen“ Wohnlage

oder eines hohen ökologischen Standards entsprechenden Hauses beim Immobilienkauf. Dabei geht eine erhöhte Nachfrage nach dem Umweltgut mit einer erhöhten Nachfrage nach einem privaten Gut einher.

Der Vorteil der *revealed* Präferenzmethoden besteht darin, dass die Personen reale Wahlen treffen bzw. reale Käufe tätigen. Allerdings weisen diese Methoden eine Reihe von Einschränkungen auf (Bennett & Blamey, 2001; Boyle, 2003): Sie erfassen nur den Gebrauchswert eines Umweltgutes und die passiven Nutzenwerte werden von diesen Methoden vollständig ausgeklammert. Bei Umweltgütern mit einem hohen passiven Nutzen wie z. B. seltene Tier- und Pflanzenarten lässt sich die Wertschätzung nicht über beobachtbares Marktverhalten erfassen, da zu diesen keine komplementären Marktgüter existieren (Adamowicz, Boxall, M. Williams & Louviere, 1998). Weitere Kritikpunkte betreffen auch die Validität der durch die *revealed* Präferenzmethoden ermittelten Zahlungsbereitschaften (Bateman et al., 2002; Pommerehne, 1987; Pommerehne & Römer, 1992).

In der Marktforschung wird die Anwendung von *revealed* Präferenzmethoden zur Ermittlung von Zahlungsbereitschaften ebenfalls kritisiert (vgl. Breidert et al., 2006 für einen Überblick). Die Preise im Handel weisen bspw. eine sehr geringe Varianz auf, so dass die darauf basierenden Schätzungen nur als eine Mindestgrenze der wahren Zahlungsbereitschaft zu interpretieren sind; Daten aus Testmärkten bzw. Testmarktsimulatoren umgehen dieses Problem durch systematische Variation der Kaufpreise, sind aber mit erheblichen Kosten verbunden und finden eher selten Anwendung im Marketingbereich. Die *revealed* Präferenzmethoden können in der Marktforschung ebenfalls nicht verwendet werden, wenn es sich um die Bewertung von einem neuen Produkt oder Service handelt, die in Planung sind und daher noch nicht existieren.

Bei Umweltgütern mit einem hohen Nicht-Gebrauchsnutzen sowie bei neuen Konsumgütern besteht die einzige Möglichkeit, deren Nutzen zu erfassen, in der Anwendung von s.g. *stated* Präferenzmethoden oder Befragungsmethoden (vgl. Bateman et al., 2002). Diese alternative Methodenklasse beansprucht, den Gesamtnutzen (Gebrauchs- und Nicht-Gebrauchsnutzen) eines (Umwelt-)Gutes zu erfassen. Die Zahlungsbereitschaft für das (Umwelt-)Gut wird auf einem konstruierten hypothetischen Markt im Rahmen von Befragungen ermittelt (R.T. Carson, 1991). Im Vergleich zu den *revealed* Präferenzmethoden wird der Nutzen aus Verhalten in hypothetischen Situationen bzw. aus Verhaltensintentionen anstatt aus realem Verhalten geschätzt (Hanley et al., 2001).

In der Umweltbewertungsforschung zählen zu den Befragungsmethoden die traditionelle Contingent Valuation Methode (CV) und die neueren Ansätze der Conjoint- und der Discrete-Choice-Analysen. Die zwei letztgenannten Verfahren werden häufig in der Umweltbewertung

unter der Bezeichnung *Choice Experiments* zusammengefasst (vgl. Hanley et al., 2001). In der Marktforschungsliteratur werden beide Methoden hingegen oft unter dem Begriff der *Conjoint Analyse* zusammengefasst (vgl. Bredert et al., 2006). Ein wesentlicher Unterschied zwischen den drei Befragungsmethoden besteht in der Abfrage bzw. Erfassung der Zahlungsbereitschaften der Personen. Der Contingent Valuation-Ansatz erfragt *direkt* die Zahlungsbereitschaft der Personen für eine Veränderung in der Qualität oder Quantität eines Umweltgutes (z.B. einer Naturlandschaft). Im Rahmen der Conjoint- und der Discrete-Choice-Analysen wird hingegen angenommen, dass sich der Nutzen eines Gutes aus dem Nutzen seiner Charakteristika zusammensetzt (Lancaster, 1966). Die Naturlandschaft z. B. wird anhand von verschiedenen Merkmalen mit verschiedenen Merkmalsausprägungen beschrieben (Anteile von Waldflächen, von Weidenflächen, Fahrradwegen und Kosten für die Realisierung). Aufgrund dieser Besonderheit sind die CA- und DC-Analyse in der Forschung auch unter der Bezeichnung „multiattributive Ansätze“ bekannt (vgl. Hanley et al., 2001). Die Zahlungsbereitschaft in den multiattributiven Ansätzen wird für einzelne Merkmale und ganzheitliche Optionen *indirekt* aus der Präferenzstärke für verschiedene Alternativen bzw. aus den Wahlen zwischen Alternativen abgeleitet (vgl. Abschnitt 2.4).

## 2.3 Die Befragungsmethoden

### 2.3.1 Contingent Valuation

Contingent Valuation ist eine Methodenklasse, die ihre Wurzeln in der Umweltökonomie als Teildisziplin der Volkswirtschaftslehre hat. Die Grundidee der CV-Methode ist einfach: Wenn ein tatsächlicher Markt für ein Umweltgut fehlt, wird ein hypothetischer Markt im Rahmen einer strukturierten Befragung konstruiert (R.T. Carson, 1991). Das Umweltgut stellt dabei eine geplante Umweltveränderung dar. Die Umweltveränderung wird anhand von einem Status-Quo-Zustand (SQ; aktueller Zustand) und einem verbesserten Umweltzustand infolge bspw. einer Naturschutzmaßnahme, die mit Kosten verbunden ist, beschrieben.

Die Wahl der Methode zum Abruf der Zahlungsbereitschaft (*elicitation method*; Frageformat) stellt eine der wichtigsten konzeptuellen Fragen bei der Gestaltung einer CV-Studie. Diesbezüglich durchlief der CV-Ansatz mehrere Entwicklungen, die zu einer Fülle an Frageformaten geführt haben. Diese lassen sich in Anlehnung an Mitchell und R.T. Carson (1989) in vier größere Kategorien unterteilen. Maßgebend dabei ist die Anzahl gestellter Fragen (eine vs. eine Reihe) und das Antwortformat (offen vs. geschlossen). Tabelle 1 enthält Beispiele und eine Zuordnung von den populärsten CV-Frageformaten zu den o.g. Kategorien.

### Frageformate

In den Anfängen der CV-Methode wurden hauptsächlich verschiedene offene Formate eingesetzt. In den offenen Formaten haben die Befragten den Geldbetrag zu finden und anzugeben, der ihren Präferenzen am besten entspricht. Die Befragten können dabei ihre Zahlungsbereitschaft im s.g. offenen Format direkt nennen (*open ended*; s. Tabelle 1) oder mithilfe der s.g. Zahlkarte aus verschiedenen Werten denjenigen wählen, der der eigenen WTP am nächsten steht (*payment card*; s. Tabelle 1).

Tabelle 1: Klassifikation der Contingent Valuation-Varianten in Abhängigkeit von Antwortformat und Anzahl der gestellten Fragen (mit Beispielen)

| Antwortformat      | Anzahl der Fragen                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                    | Eine Frage                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | Mehrere Fragen                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <i>offen</i>       | <p><b>Open Ended-Format:</b><br/>Wie viel sind Sie bereit maximal pro Jahr an erhöhten Strompreisen zu bezahlen, damit der Anteil des Öko-Stroms in der Stromversorgung um 50% steigt?</p> <p><b>Payment Card-Format:</b><br/>Wie viel sind Sie bereit maximal pro Jahr an erhöhten Strompreisen zu bezahlen, damit der Anteil des Öko-Stroms in der Stromversorgung um 50% steigt? Vermerken Sie den Geldbetrag, der am besten Ihrer maximalen Zahlungsbereitschaft entspricht:<br/><i>0 €, 1 €, 2 €, ..., 10 €, 15 €, ..., &gt; 200 €</i></p> | <p><b>Bidding Game-Format:</b><br/>Würden Sie 20€ mehr im Jahr an erhöhten Strompreisen bezahlen für den Anstieg des Öko-Strom-Anteils in der Stromversorgung um 50%?<br/>→ <b>JA:</b> Würden Sie 30€ mehr bezahlen?<br/>(→ <b>JA:</b> 40€?   <b>NEIN:</b> 25€?)<br/>...etc. Wie viel genau?<br/>→ <b>NEIN:</b> Würden Sie 10€ mehr bezahlen?<br/>(→ <b>JA:</b> 15€?   <b>NEIN:</b> 5€?)<br/>...etc. Wie viel genau?</p>                                       |
| <i>geschlossen</i> | <p><b>Closed Ended-Format:</b><br/>Würden Sie 20<sup>a</sup> € mehr im Jahr an erhöhten Strompreisen bezahlen für den Anstieg des Öko-Strom-Anteils in der Stromversorgung um 50%?</p> <p style="text-align: center;"><b>JA</b>                      <b>NEIN</b></p>                                                                                                                                                                                                                                                                            | <p><b>Double Bounded-Format:</b><br/>Würden Sie 20<sup>a</sup> € mehr im Jahr an erhöhten Strompreisen bezahlen für den Anstieg des Öko-Strom-Anteils in der Stromversorgung um 50%?<br/>→ <b>JA:</b> Würden Sie 30 € mehr bezahlen?<br/><span style="float: right;"><b>JA</b>                      <b>NEIN</b></span><br/>→ <b>NEIN:</b> Würden Sie 10 € mehr bezahlen?<br/><span style="float: right;"><b>JA</b>                      <b>NEIN</b></span></p> |

Anmerkungen. <sup>a</sup> die Beträge werden über die Fragen variiert.

Ein iteratives offenes Fragformat ist das s.g. *bidding game*. Bei diesem entscheiden die Personen, ob sie einer Maßnahme zu einem bestimmten Preis zustimmen würden. Bei einer Zustimmung wird der Preis solange angehoben, bis die Person „Nein“ sagt. Bei einer Ablehnung des ursprünglichen Preises, wird dieser solange gesenkt bis die Person zustimmt. Zum Schluss

werden die Personen direkt nach ihrer WTP gefragt (vgl. Tabelle 1). Obwohl die offenen Formate leicht anzuwenden und auszuwerten sind, gibt es Zweifel an der Zuverlässigkeit der daraus abgeleiteten WTP-Schätzungen (Arrow et al., 1993; Diamond & Hausman, 1994).

Das offene Frageformat führt z. B. oft zu fehlenden Angaben oder zu hohen Anteilen an Null-Antworten sowie zu wenigen sehr hohen Werten. Diese Inkonsistenzen werden als Indiz für verschiedene Antwortverzerrungen interpretiert. Dazu zählen kognitive Überforderung durch die unvertraute Aufgabe (Bateman, Langford, R.K. Turner, K.G. Willis & Garrod, 1995; Schkade & Payne, 1994), moralische Überlegungen (Kahnemann & Ritov, 1994) und strategisches Verhalten (Hanemann, 1994; Hoehn & Randall, 1987; Kealy & R.W. Turner, 1993).

Die *bidding game*- und *payment card*-Formate werden ebenfalls wegen möglicher Antwortverzerrungen kritisiert. Beim *bidding game* zeigt sich bspw., dass die Höhe der ermittelten Zahlungsbereitschaft mit der Höhe des vorgegebenen Preises für die Umweltmaßnahme korreliert (vgl. Boyle, Bishop & Welsh, 1985). Diese Inkonsistenz ist in der Literatur unter der Bezeichnung *starting point bias* bekannt und stellt eine Art Ankereffekt dar (für eine ausführliche Darstellung zum Ankereffekt vgl. Strack & Mussweiler, 1997; Tversky & Kahneman, 1974). Die *payment card*-Methode unterliegt einem Spannbreite-Effekt, bei dem die Höhe der ermittelten WTP durch die vorgegebenen Preisspannen beeinflusst wird (Mitchell & Carson, 1989). Die ermittelten Inkonsistenzen in den offenen CV-Frageformaten veranlassten die NOAA-Kommission, ein speziell zur Beurteilung der Güte der CV-Methode einberufenes Expertengremium, diese Formate als unzuverlässig einzuschätzen (Arrow et al., 1993).

Obwohl in neueren Arbeiten die Verbannung der offenen Formate als voreilig angesehen wird (D. Green et al., 1998; Mitani & Flores, 2009; Ready, Buzby & Hu, 1996; Spash, ten Bosch, Westmacott & Ruitenbeek, 2000), führte die Empfehlung der NOAA-Kommission zur favorisierten Anwendung des s.g. geschlossenen CV-Frageformates. Dieses Format ist im englischsprachigen Raum unter verschiedenen Bezeichnungen bekannt - unter anderem als *closed ended*- (CE), *dichotomous choice*-, *referendum*-, *take-or-leave-it*- oder *single bounded*-Format (Herriges & Shogren, 1996). In dieser Form der CV-Methode entscheiden sich die Befragten ähnlich einer Marktsituation, ob sie eine umschriebene Naturschutzmaßnahme zu einem bestimmten Preis akzeptieren oder ablehnen würden (vgl. Tabelle 1).

Eine Zustimmung im geschlossenen Format (CE) ist ein Indiz dafür, dass die WTP des Befragten höher oder gleich dem präsentierten Preis ist. Eine Ablehnung impliziert, dass die WTP des Befragten unter dem Preis liegt. Die Preise werden über die Befragten hinweg variiert. Für die Berechnung der WTP ergeben sich aus CE somit keine individuellen Werte, sondern lediglich Intervalle. Die Auswertung erfolgt auf Populationsebene mithilfe binärlogistischer Modelle (s. Alberini, 1995; Hanemann & Kanninen, 1999).

Der CE-Frage können weitere geschlossene Fragen folgen, um der wahren WTP der Befragten näher zu kommen und die statistische Effizienz der Methode zu steigern (Hanemann, Loomis & Kanninen, 1991). In Abhängigkeit davon, ob es sich um eine, zwei oder mehrere anschließende Fragen handelt, werden die Frageformate entsprechend *double-*, *triple-* bzw. *multiple-bounded* genannt (Bateman, Langford, A.P. Jones & Kerr, 2001; Welsh & Poe, 1998). Ein Beispiel für das *double-bounded* CV-Format kann Tabelle 1 entnommen werden.

Wesentliche Nachteile des CE-Formates gegenüber dem offenen Format sind die geringere statistische Effizienz, die Anwendung komplexer Designprozeduren (z. B. bei der Gestaltung der Preisvariable), die komplexere Auswertung der Daten aufgrund ihres binären Charakters, keine individuellen Schätzungen sondern Schätzungen für die Gesamtpopulation sowie deutliche Variation der WTP-Schätzungen in Abhängigkeit von der Modell- und Prozedurwahl (Hanemann, 1989; Johansson, Kriström & Mäler, 1989). Im Vergleich zu OE hat CE den Vorteil, einer vertrauten realen Marktsituation zu ähneln und geringere kognitive Anforderungen an die Befragten zu stellen. Als ein grundlegender Vorteil von CE wird auch seine Anreizkompatibilität hervorgehoben (Arrow et al., 1993; R.T. Carson & Groves, 2007; Hoehn & Randall, 1987). Eine Methode wird dann als anreizkompatibel bezeichnet, wenn sie *theoretisch* die wahren Präferenzen und Zahlungsbereitschaften der Personen hervorruft (s. Kapitel 5).

Obwohl das geschlossene Format eine Reihe von praktischen und theoretischen Vorteilen aufweist, sind seine Antworten ebenfalls für verschiedene Verzerrungen anfällig. Es führt z. B. meist zu höheren Durchschnittswerten der Zahlungsbereitschaft im Vergleich zum offenen Format (Cameron, Poe, Ethier & Schulze, 2002; Stevens, Belkner, Dennis, Kittredge & C.E. Willis, 2000). Als mögliche Ursache dafür wurden falsche Zustimmungen zu dem in CE vorgegeben Preis diskutiert, der s.g. „Ja-Sagen“-Effekt (Blamey, Bennett & Morrison, 1999; Boyle, McDonald, Cheng & McCollum, 1998; Holmes & Kramer, 1995), oder auch Ankereffekte genannt (Fryklblom & Shogren, 2000; D. Green et al. 1998; Holmes & Kramer, 1995; O'Connor, Johannesson & Johansson, 1999). Weiterhin konnte in Experimenten aufgezeigt werden, dass die CE-Antworten in einer Befragung sich signifikant von den Antworten und von den Zahlungen der gleichen Personen unterscheiden, wenn diese aufgefordert werden tatsächlich für das Gut zu bezahlen. Diese s.g. hypothetische Verzerrung wird in Abschnitten 2.5.2 und 3.1 weiter ausgeführt.

Die Schätzungen aus den *double-* und *multiple-bounded* Frageformaten weisen ebenfalls verschiedene Inkonsistenzen auf. Sie fallen z.B. geringer aus als die Schätzungen aus CE, die als anreizkompatibel gilt (Bateman et al., 2001; Cameron & Quiggin, 1994) und können strategischem Verhalten, Framing- und Anker-Effekten unterliegen (s. Übersicht in DeShazo, 2002).

*Bereitstellungs- und Zahlungsregeln*

Neben dem Frageformat stellen die Bereitstellungs- und Zahlungsmechanismen für das Umweltgut wichtige Kernelemente einer CV-Befragung dar (R.T. Carson, 2000). Der Bereitstellungsmechanismus spezifiziert den institutionellen Kontext der Bereitstellung (Soll z. B. die Regierung, die Kommune oder ein Treuhänderfond ein Umweltprojekt realisieren?) und die Regel für die Bereitstellung des Umweltgutes. Unter der letzteren wird das Prinzip aufgefasst, nach dem die individuellen Antworten bzw. Präferenzen der Befragten aggregiert und in eine kollektive Entscheidung über die Bereitstellung oder Nicht-Bereitstellung des Umweltgutes überführt werden. Vereinfacht ausgedrückt, beschreibt die Bereitstellungsregel, wie die Entscheidung über die Bereitstellung des Gutes mit den Antworten der Befragten zusammenhängt (Polome, 2003; L.O. Taylor et al., 2010). Die Pluralitäts- und die Majoritätsregel sind dabei die meistverwendeten Mechanismen (Arrow et al., 1993).

Der Zahlungsmechanismus beschreibt den Kontext, in dem die Person im Falle der Bereitstellung für das Gut zu bezahlen hat (Cummings et al, 1986; Morrison, Blamey & Bennett, 2000). In diesem wird zum einen die Zahlungsregel spezifiziert (Wie viel Geld ist zu bezahlen?) und zum anderen das Zahlungsverkehrsmittel (In welcher Form und wie oft ist die Bezahlung zu entrichten?). Die Zahlung für das Umweltgut kann an der Antwort der Person in der CV-Befragung direkt *gekoppelt* sein - die Person bezahlt den Betrag, dem sie z. B. im geschlossenen Format zugestimmt hat bzw. im offenen Format selbst angegeben hat. Bei *entkoppelter* Zahlung bezahlt die Person z. B. einen festen Betrag, dessen Höhe nicht mit ihrer Antwort zusammenhängt (vgl. D. Green et al. 1998). Die Zahlung kann durch den Einsatz verschiedener Verkehrsmittel erfolgen wie Spenden, Einkommens-, Umsatz-, Vermögenssteuer, Zahlungen in Treuhänderfonds, Eintrittsgebühren, Änderungen in Produktpreisen (Garrod & K.G. Willis, 1999). Allgemein lassen sich die Zahlungsverkehrsmittel in freiwillige (z. B. Spende an einer Naturschutzorganisation, das Beziehen von Öko-Strom im Haushalt) und verpflichtende Zahlungen (z. B. Erhöhung der KFZ-Steuer, der Abwassergebühren, der Strompreise) einteilen (Bateman et al., 2002). Die verschiedenen Bereitstellungs- und Zahlungsregeln können unterschiedliche Anreize für strategisches oder wahrheitsgemäßes Antwortverhalten bei den Befragten schaffen (D. Green et al., 1998; vgl. Abschnitt 5.1.2).

Zusammenfassend ist die CV-Methode eine lange Entwicklung durchlaufen in der Bemühung, die wahren Zahlungsbereitschaften der Befragten zu erfassen. Diese Entwicklung kann als ein Lernprozess aufgefasst werden, bei dem jedes weitere CV-Format zum Abruf von Zahlungsbereitschaften die Nachteile seiner Vorgänger zu kompensieren und Verzerrungen zu vermeiden versuchte. Die Frage, welches CV-Format am besten geeignet ist, die wahre WTP der Befragten zu erfassen, ist jedoch bis heute ungelöst geblieben. Inkonsistenzen wie die hypothetische Verzerrung, Anker- und Framingeffekte sowie die Möglichkeit, durch strategische Antworten, die Ergebnisse von CV-Befragungen zu beeinflussen, sind einige der wich-

tigsten Fehlerquellen, die die Validität von CV beeinträchtigen können. Vor diesem Hintergrund ist das gestiegene Interesse in der aktuellen Umweltbewertungsforschung an der Conjoint- und der Discrete Choice-Analyse als Reaktion auf die scharfe Kritik an die CV-Methode in den 90ern Jahren (vgl. Hausman, 1993) zu verstehen.

### 2.3.2 Die multiattributiven Ansätze

Die Methoden der Conjoint- und der Discrete Choice-Analysen haben ihren Ursprung unter anderem in der mathematischen Psychologie (Luce & Tukey, 1964; Thurstone, 1927). Sie finden seit längerem eine breite Anwendung in Marktforschung und Transportwirtschaft (P.E. Green & Srinivasan, 1978; Hensher, 1994; Louviere, 1994) und wurden Mitte der 90ern zum ersten Mal in der Umweltbewertung implementiert (Adamowicz, Louviere & M. Williams, 1994). CA und DC sind Methoden, mit denen in erster Linie die Präferenzen für Güter und Services modelliert werden. Unter Präferenzen sind dabei vergleichende Urteile zu verstehen, die unter bestimmten Voraussetzungen durch eine numerische Skala repräsentiert bzw. als Nutzen bezeichnet werden können (vgl. Ben-Akiva et al., 1999). CA und DC basieren weiterhin auf der Annahme, dass Güter und Services als Bündel von Eigenschaften oder Attributen wahrgenommen werden. Der Gesamtnutzen eines Gutes ergibt sich aus den Nutzenwerten der einzelnen Eigenschaften bzw. ihrer Ausprägungen (Lancaster, 1966). Ein Naturpark lässt sich z. B. anhand der Tier- und Pflanzenvielfalt (große vs. geringe Artenvielfalt) sowie der Freizeitmöglichkeiten dort (Wandern vs. Wassersport) charakterisieren. In der Regel wird angenommen, dass der Gesamtnutzen sich additiv aus den s.g. Teilnutzenwerten der einzelnen Eigenschaften bzw. ihren Ausprägungen zusammensetzt (Hanley et al., 2001). Im Rahmen von CA und DC werden die Personen mit mehreren alternativen Beschreibungen des Gutes (die s.g. Alternativen/ Optionen/ Stimuli/ Profile/ Szenarien) konfrontiert, die sich in ihren Merkmalsausprägungen unterscheiden. Die Personen sollen in CA ihre Präferenzstärke für die verschiedenen Alternativen angeben z.B. in Form von Ratings, Rankings oder Paarvergleiche. In DC treffen sie aus den präsentierten Alternativen eine Wahl. Aus den Präferenzen bzw. der Wahlen der Befragten für die ganzheitlichen Alternativen wird auf die Teilnutzenwerte der einzelnen Merkmalsausprägungen geschlossen. Indem eine Preis- oder Kostenvariable als Merkmal eingeschlossen wird, lassen sich die Zahlungsbereitschaften für einzelne Merkmale und Alternativen ableiten. DC und CA erfordern ein komplexes methodenspezifisches Studiendesign, das in Gustafsson, Herrmann und F. Huber (2003) näher beschrieben wurde. An dieser Stelle erfolgt nur ein kurzer Umriss der Schätzung von Präferenzen aus beiden Methoden, da diese die Grundlage für die Berechnung von WTP dienen und im Fokus der vorliegenden Arbeit stehen.

*Spezifikation der Nutzenfunktion und Schätzung der Präferenzparameter in den multiattributiven Ansätzen*

Die Conjoint Analyse (CA) bedient sich einer deterministischen Nutzenfunktion. Dabei werden die Präferenzurteile der Befragten (Ratings oder Rankings) für verschiedene Alternativen direkt in eine Nutzenskala transformiert. In diesem Zusammenhang wird angenommen, dass die indirekte Nutzenfunktion  $U$  und die beobachtbaren Präferenzen  $V$  (in Form von Ratings/Rankings-Angaben) von Person  $i$  für Alternative  $j$  direkt mithilfe einer Transformationsfunktion  $\phi$  verbunden sind (vgl. Hanley et al., 2001):

$$U_{ij} = \phi[V_{ij}(X_{ij})] \quad (1)$$

Die beobachtbare Präferenz ist ihrerseits eine Funktion der Merkmale  $X_{ij}$  der Optionen. Die Transformationsfunktion kann verschiedene Formen annehmen und korrespondiert mit verschiedenen CA-Modellen (vgl. Backhaus et al., 2000). Im s.g. Teilnutzenwertmodell wird für jede Ausprägung eines Merkmals einen separaten Teilnutzenwert geschätzt. Dabei werden die  $l$  Eigenschaftsausprägungen eines Merkmals  $a$  in  $(l-1)$  Dummy-Variablen umkodiert. Die am wenigsten wünschenswerte Eigenschaftsausprägung bzw. der Status-Quo-Zustand dient als Referenzgröße, um die anschließende Interpretation der Ergebnisse zu erleichtern. Weiterhin wird von einer additiven Nutzenfunktion ausgegangen (Lilien, Kotler & Moorthy, 1992). Das Teilnutzenwertmodell kann für beliebige Funktionsverläufe definiert werden, so dass verschiedene Modelle wie das Idealpunkt- oder das lineare Vektormodell als Spezialfälle des Teilnutzenwertmodells aufgefasst werden können (Gutsche, 1995).

Die Schätzung von Teilnutzenwerten in CA erfolgt in Abhängigkeit vom Meßniveau der Präferenzurteile. Für die Auswertung von ordinalskalierten Rangurteilen aus Ranking-Aufgaben kann MONANOVA eingesetzt werden (Kruskal, 1965). Die intervallskalierten Ratings werden mithilfe der OLS-Regression analysiert (Darmon & Rouzies, 1994). Da die OLS als robustes Schätzverfahren gilt, und von diverser Software unterstützt wird, sind viele Forscher im Marketingbereich dazu übergegangen, Ranking-Daten auf intervallskaliertes Niveau aufzuwerten und für ihre Auswertung die OLS-Regression anzuwenden. Mehrere Studien konnten aufzeigen, dass die Schätzgenauigkeit dadurch nicht negativ beeinflusst wird (Carmone, P.E. Green & Jain, 1978; Wittnik & Cattin, 1981). In der Umweltbewertung werden zur Auswertung der Ranking-Methode ähnlich der DC vorwiegend Wahlmodelle eingesetzt.

Die Discrete Choice-Analyse (DC) stellt eine Integration verschiedener theoretischer Bereiche dar, unter anderem Lancaster's Theorie, Wohlfahrts- und Konsumententheorie und experimentelle Design-Theorie (Hanley et al., 2001; Lancsar & Savage, 2004). Die Modellierung von Wahlen aus DC beruht wie die Analyse der Antworten aus dem geschlossenen CV-Format auf die Random-Utility-Theorie (RUT; Zufallsnutzentheorie; McFadden, 1973).

In RUT wird angenommen, dass die Person bei mehreren Alternativen sich stets für diejenige entscheidet, die ihr den größten Nutzen bringt. Im Gegensatz zu CA entspricht im RUT-Modell der beobachtbare Nutzen nicht dem wahren Nutzen, da die Wahlen einer Person durch verschiedenen nichtbeobachtbare Faktoren beeinflusst werden können. Im Vergleich zu CA wird der wahre Nutzen einer Option  $j$  für Person  $i$  ( $U_{ij}$ ) als eine latente Variable aufgefasst, die sich in eine deterministische, beobachtbare Nutzenkomponente ( $V_{ij}$ ) und eine stochastische Nutzenkomponente ( $e_{ij}$ ) zerlegen lässt (Train, 2003):

$$U_{ij} = V_{ij} + e_{ij} \quad (2)$$

Die stochastische Fehlerkomponente  $e_{ij}$  repräsentiert alle Einflüsse auf die Wahl der Person, die nicht beobachtbar sind wie bspw. nicht erfasste Merkmale des Gutes, nicht beobachtbare Präferenzvariabilität, Spezifikationsfehler im Modell etc. (Ben-Akiva & Lerman, 1985). Die deterministische, oder beobachtbare, Komponente wird im RUT-Modell somit als eine Funktion von Merkmalen des Gutes, der Person etc. dargestellt, die dem Forscher bekannt sind bzw. die von ihm erfasst wurden:

$$V_{ij} = X'_{ij}\beta \quad (3)$$

wobei  $X_{ij}$  ' der Vektor der Merkmale von Gut und Person bzw. der dummy-kodierten Merkmalsausprägungen darstellt (vgl. Train, 2003). Wenn die Wahl einer Option (z. B. den Besuch eines Nationalparks) allein anhand der Merkmale vom Gut modelliert wird (z. B. Artenvielfalt, Gewässer, Freizeitmöglichkeiten, Eintrittspreis etc.), lässt sich die deterministische Komponente  $V_j$  als eine bspw. lineare Funktion dieser Merkmale auffassen:

$$V_j = \beta_1 \text{Preis}_{j1} + \beta_2 \text{Artenvielfalt}_{j2} + \beta_3 \text{Gewässer}_{j3} + \dots \quad (4)$$

Die beobachtbaren Wahlen in DC sind im Vergleich zu den Ratings und Rankings in CA lediglich als Indikatoren für den wahren Nutzen aufzufassen. Aufgrund der Annahme einer zufälligen Nutzenkomponente ermöglichen RUT-Modelle nur probabilistische Aussagen über das Wahlverhalten der Personen. Ausgehend von einem nutzenmaximierenden Wahlverhalten wählt Person  $i$  die Alternative  $q$  aus einem Set mit  $J$  Alternativen mit einer Wahrscheinlichkeit von  $p_{iq}$ , wenn Alternative  $q$  ihr den größten Nutzen stiftet:

$$\begin{aligned} p_{iq} &= P(U_{iq} > U_{ij}) = P[(V_{iq} + e_{iq}) > (V_{ij} + e_{ij})] \\ &= P[(V_{iq} - V_{ij}) > (e_{ij} - e_{iq})], \quad \forall j \neq q \end{aligned} \quad (5)$$

wobei  $p_{iq}$  eine Variable ist, die das Ergebnis der Wahl darstellt (Train, 2003). Wie Formel (5) zeigt, wird in DC die Auswahlentscheidung nicht anhand des absoluten Nutzens der Alternativen modelliert, sondern anhand der Nutzendifferenzen der Alternativen. Um die Wahlwahrscheinlichkeit für eine Alternative zu berechnen, gilt es die Verteilung der Fehlerterme  $e$  zu kennen bzw. sind Annahmen über deren Verteilungen zu treffen. Unterschiedliche Annahmen

über die Struktur und die Verteilung der Fehlerkomponente führen in DC zu unterschiedlichen Modellen (s. Train, 2003 für eine Übersicht).

**Das Multinomiale-Logit-Modell:** Die Annahme über unabhängige und identisch verteilte (*iid*) Fehler in DC, die der Gumbleverteilung folgen, impliziert, dass die Differenzen der Fehlerterme in Gleichung 5 logistisch verteilt sind. Daraus resultiert im Fall von zwei Alternativen das binomiale Logit-Modell (Hanemann, 1984) und bei mehr als zwei Alternativen das Konditionale- bzw. das Multinomiale-Logit-Modell (MNL-Modell; McFadden, 1973).

Die Wahlwahrscheinlichkeit für Alternative  $q$  aus einem Wahlset mit  $J$  Alternativen kann unter Berücksichtigung der Merkmale der Alternative (vgl. Gleichung 3) im MNL-Modell wie folgend berechnet werden:

$$P_{iq} = \frac{\exp(\lambda V_{iq})}{\sum_{j=1}^J \exp(\lambda V_{ij})} = \frac{\exp(\lambda \beta x'_{iq})}{\sum_{j=1}^J \exp(\lambda \beta x'_{ij})} \quad (6)$$

wobei  $\lambda$  ein positiver Skalenparameter darstellt, der invers proportional zur Fehlervarianz  $\sigma_e$  der Zufallskomponente ist (vgl. Swait & Louviere, 1993). Innerhalb eines Modells kann der Skalenfaktor  $\lambda$  nicht bestimmt werden und wird auf 1 fixiert (Ben-Akiva & Lerman, 1985).

Die Schätzung der Präferenzparameter  $\beta$  in Gleichung 6 erfolgt mithilfe des Maximum-Likelihood-Schätzverfahrens. Bei dieser Methode werden die Parameter aus den Daten so geschätzt, dass die Wahrscheinlichkeit für die tatsächlich beobachteten Wahlen maximiert wird. Die geschätzten Präferenzparameter sind jedoch stets mit dem Skalenparameter konfundiert, so dass diese nicht als  $\beta$ , sondern als  $\beta/\sigma_e$  aufzufassen sind (Train, 2003). Der direkte Vergleich der Parameterschätzungen zwischen verschiedenen MNL-Modellen ist daher nicht möglich; die Ermittlung der Skalenfaktoren in verschiedenen Modellen kann jedoch über das Verhältnis der Skalenparameter erfolgen (vgl. Swait & Louviere, 1993).

Eine Implikation aus Gleichung (6) ist die s.g. IIA-Annahme (Independence of Irrelevant Alternatives; Luce, 1959). Diese Eigenschaft resultiert aus der Annahme unabhängiger Fehlerterme (*iid*) verschiedener Alternativen im Choice-Set und besagt, dass das Verhältnis der Wahlwahrscheinlichkeit von zwei Optionen unverändert bleiben soll, wenn aus dem Wahlset Alternativen entfernt oder in das Wahlset andere Alternativen eingeführt werden (vgl. Train, 2003). Zur Überprüfung der IIA-Annahme liegt z. B. der Hausman-McFadden-Test vor (Hausman & McFadden, 1984). Eine andere Möglichkeit mit dieser restriktiven Annahme umzugehen bzw. die *iid*-Annahme zu lockern, besteht in der Anwendung von alternativen Modellen wie bspw. Nested-Logit- (Ryan & Skatun, 2004), Latent-Class- (Swait & Adamowicz, 2001) oder Mixed-Logit-Modelle (McFadden & Train, 2000; s.u.).

**Das Mixed-Logit-Modell:** Das Mixed-Logit-Modell (MXL) ist eine flexible Alternative für die ökonometrische Modellierung von diskreten Wahlen. Es wird auch noch als random pa-

parameter-Modell bezeichnet, denn die geschätzten Parameter sind im Kontrast zu den Parametern des MNL-Modells nicht fixiert, sondern stellen Zufallsvariablen dar (Train, 2003). Jedes DC-Modell kann unter bestimmten Bedingungen als ein Spezialfall des MXL-Modells dargestellt werden (McFadden & Train, 2000).

Im MXL-Modell wird angenommen, dass die Präferenzparameter in der Population mit einer bekannten Verteilung mit der Dichtefunktion  $f(\beta | \theta)$  variieren (Train, 1998). Der latente Nutzen einer Option  $j$  für die Person  $i$  wird als die Summe von 1) dem Mittelwert  $b$  der Präferenzen in der Population; 2) einem Parameter  $\mu_i$ , der der individuellen Abweichung im Geschmack vom Mittelwert der Population Rechnung trägt, und 3) einer nicht beobachtbaren *iid*-verteilten Fehlerkomponente  $e_{ij}$  aufgefasst.

Der Nutzen aus Option  $j$  für Person  $i$  lässt sich im MXL-Modell formell darstellen als:

$$U_{ij} = X'_{ij}b + (X'_{ij}\mu_i + e_{ij}) = X'_{ij}(b + \mu_i) + e_{ij} = \beta_i X'_{ij} + e_{ij} \quad (7)$$

Dabei stellt  $X'_{ij}b$  die deterministische Komponente im Modell dar, und  $X'_{ij}\mu_i + e_{ij}$  ist die stochastische Komponente. Die stochastische Komponente ist korreliert zwischen den Alternativen, so dass im MXL-Modell die IIA-Annahme nicht mehr erforderlich ist und die wesentlichen Einschränkungen des MNL-Modells dadurch aufgehoben werden (Alpizar et al. 2003; Train, 2003).

Die im MNL-Modell fixierten  $\beta$  variieren im MXL-Modell zwischen den Personen  $\beta_i = b + \mu_i$  und tragen somit Rechnung der Präferenzheterogenität in der Population. Die Werte von  $\beta_i$  hängen von den Parametern der angenommenen Verteilung  $f(\beta | \theta)$  ab. Die Normal-, Lognormal- und die Dreiecksverteilung gehören zu den üblichen Verteilungsformen, die Anwendung in der Forschungsliteratur finden (Hensher & Greene, 2003; Train, 2003). Eine Lognormalverteilung wird dann unterstellt, wenn die zu schätzenden Parameter ein bestimmtes (positives oder negatives) Vorzeichen aufweisen sollen (Hensher & Greene, 2003). Ein Beispiel dafür wäre die Preisvariable, die für gewöhnlich einen negativen Nutzen hat.

Für den Parameter  $e_{ij}$  wird im MXL-Modell wie im MNL-Modell eine Gumbleverteilung angenommen (Hensher & Greene, 2003). Die Auswahlwahrscheinlichkeit einer Alternative  $q$  aus einem Set mit  $J$  Alternativen ergibt sich im MXL-Modell als Integral der  $\beta_i$  konditionalen Wahlwahrscheinlichkeiten aus dem MNL-Modell (vgl. Gleichung 6) gewichtet mit den Werten der Dichtefunktion (Train, 2003):

$$P_{iq} = \int \frac{\exp(\beta_i X'_{iq})}{\sum \exp(\beta_i X'_{ij})} f(\beta) d\beta \quad (8)$$

Im MXL-Modell werden somit die Auswahlentscheidungen als Integral über alle möglichen Werte von  $\beta$  modelliert. Die Schätzung der Präferenzparameter, ihrer Varianzen und Kovarianzen erfolgt nicht analytisch, sondern mithilfe von Simulationsmethoden. Hierzu können die

simulierte Maximum-Likelihood-Methode (s. Bhat, 2001; Train 2003) oder das Hierarchische-Bayes-Schätzverfahren herangezogen werden (s. Allenby & Rossi, 1998). Beide Simulationmethoden sind in ihren Schätzungen ähnlich und ermöglichen sowohl die Ermittlung von Mittelwerten und Standardabweichungen der Parameter als auch die Schätzung individueller Nutzenwerte (Huber & Train, 2001). Somit lassen sich aus dem MXL-Modell im Vergleich zum MNL-Modell nicht nur populationsspezifische, sondern auch individuelle Nutzenwerte und Zahlungsbereitschaften schätzen.

**Das Rank-Ordered-Logit-Modell:** Die Ranking-Aufgabe in CA kann als eine Reihe von Wahlen aufgefasst werden. Bei einem stufenweise ablaufenden Prozess wählt die befragte Person aus einem Wahlset mit  $J$  Alternativen zuerst die Alternative  $i1$  aus, die ihr den größten Nutzen stiftet. Nachdem diese Alternative aus dem Wahlset entfernt wurde, wählt die Person aus dem Wahlset mit  $J-1$  Alternativen die Option  $i2$  mit dem größten Nutzen aus usw. Die Antworten der Befragten in der CA-Ranking-Aufgabe lassen sich somit basierend auf RUT mithilfe des Rank-Order-Logit-Modells auswerten (Beggs, Cardell & Hausman, 1981; Chapman & Staelin, 1982, Hausman & Ruud, 1987). Das ROL-Modell ist die bevorzugte Analyse-methode in der Umweltbewertungsforschung (vgl. Hanley et al., 2001), so dass es in dieser Arbeit verwendet wird.

Das ROL-Modell kann als eine Reihe von MNL-Modellen aufgefasst werden. Beggs et al. (1981) zeigen auf, dass die Wahrscheinlichkeit für eine Rangreihe als das Produkt von  $J-1$  konditionalen Wahrscheinlichkeiten (vgl. Gleichung 6) gegeben ist, wenn  $J$  die Anzahl der Alternativen ist, die es zu ranken gilt. Die Wahrscheinlichkeit, die Alternativen in einer bestimmten Art und Weise zu ranken, ist formell gegeben durch:

$$P(U_{i1} > U_{i2} > \dots > U_{ij}) = \prod_{q=1}^{J-1} \left[ \frac{\exp(\lambda b' X_{iq})}{\sum_{j=1}^J \exp(\lambda b' X_{ij})} \right] \quad (9)$$

Die Information über die meistpräferierte Alternative in einer Ranking-Aufgabe reicht aus, um die Parameter der Nutzenfunktion zu schätzen. Die zusätzliche Information, die die Berücksichtigung weiterer Alternativen mit sich bringt, führt allerdings zu effizienteren Schätzungen (Chapman & Staelin, 1982). So wie das MNL-Modell ist das ROL-Modell an der IAA-Annahme gebunden. Diese gewährleistet, dass sich die sukzessiven konditionalen logistischen Wahrscheinlichkeiten im ROL-Modell multiplizieren lassen, um die Wahrscheinlichkeit für die Rangreihe zu berechnen. Das ROL-Modell ist weniger robust gegen Fehlspezifikationen, da die IAA-Annahme in allen  $J-1$  Wahlsets vorausgesetzt wird. Das MXL-Modell kann ebenfalls bei der Analyse von Ranking-Aufgaben verwendet werden. Die Modellspezifikation sowie Berechnungsdetails können Layton (2000) entnommen werden.

*Stellenwert der multiattributiven Ansätze in der Umweltbewertung*

Die Anwendung der metrischen Conjoint-Analyse in der Umweltbewertung ist im Vergleich zu CV und DC eher begrenzt und hängt mit der eingeschränkten Möglichkeit zusammen, wohlfahrtstheoretisch-konsistente Werte zu ermitteln (Hanley, R.E. Wright & Adamowicz, 1998; Louviere, 2001). In CA werden zwar die Nutzenfunktionen der Befragten bestimmt, im Präferenzmodell sind jedoch keine Annahmen über das Entscheidungsverhalten der Befragten verankert. Diese müssen vom Forscher im Nachhinein festgelegt werden. Dabei sind unterschiedliche Annahmen über den Zusammenhang zwischen Teilnutzenwerten und realen Wahlen möglich - bei der s.g. First-Choice-Regel wird z. B. unterstellt, dass der Befragte stets die Alternative mit dem höchsten Gesamtnutzen wählen würde. Im Vergleich zu CA entfällt in DC dieser Schritt, da diese unmittelbar die Wahlwahrscheinlichkeit erfasst.

Ein weiteres Problem in CA ist die vorausgesetzte Vergleichbarkeit der Präferenzurteile zwischen verschiedenen Personen. Die OLS-Regression liefert personenspezifische Schätzungen der Nutzenfunktion. Die Teilnutzenwerte unterschiedlicher Personen sind nicht vergleichbar, denn die Befragten können bei der Bewertung der Alternativen unterschiedliche subjektive Skalen verwenden (Backhaus et al., 2006). Bei der Ermittlung der Präferenzen von Kundensegmenten oder der Gesamtpopulation werden die Schätzungen aus CA für gewöhnlich aggregiert. Die Aggregation ist an bestimmte Annahmen gebunden, die zu unterschiedlichen Schätzungen führen können (Hahn, 1997). Weiterhin impliziert die OLS-Regression die Annahme kardinaler/intervallskaliertter Präferenzurteile - eine Annahme, die der modernen Auffassung von Präferenzen als ordinalskaliert widerspricht (Bateman et al., 2002).

Die o.g. Besonderheiten im Präferenzmodell und in der Auswertung der Conjoint-Daten führen zu einer relativ geringen Anwendung der Methode in der monetären Umweltbewertung im Kontrast zur großen Verbreitung dieser Methode im Marketingbereich (Hanley et al., 2001). Lediglich die Ranking-Methode erfreut sich seit den 80ern einer breiteren Beliebtheit und Anwendung im Umweltbereich (z. B. Beggs et al., 1981; Foster & Mourato, 2000; Garrod & K.G. Willis, 1997, 1998). Dies kann mit der Anwendung von ökonometrischen Schätzmodellen wie Rank-Ordered-Logit- und Mixed-Logit-Modelle erklärt werden, die auf der Zufallsnutzentheorie basieren.

Ähnlich CA erfordert die DC-Methode von den Befragten, Änderungen in Merkmalsausprägungen gegen die Kosten für diese Änderungen abzuwägen. Ein entscheidender Vorteil von DC gegenüber CA ist die Ermittlung theoriekonsistenter monetärer Werte (s.o.). Zum einen stimmt das ökonometrische Modell in der DC-Analyse mit der Theorie rationaler und probabilistischer Wahlen überein. Zum anderen hat die Person in DC stets die Möglichkeit, eine Status-Quo-Alternative zu wählen bzw. auf den Konsum zu verzichten. Dadurch werden die Nutzen zwischen unterschiedlichen Wahlsets in DC einheitlich skaliert. Neben den o.g. me-

thodischen Vorteilen wird auch die Realitätsnähe der DC-Aufgaben hervorgehoben (Bateman et al., 2002). Anstatt Produkte zu raten oder zu ranken, haben die Personen mehr Erfahrung und sind vertrauter mit Marktsituationen, wo sie entscheiden müssen, welches Produkt sie kaufen würden. Kahneman und Sugden (2005) schließen daher, dass Präferenzen durch die Auswahl aus Alternativen bzw. mithilfe der DC-Analyse am besten erfassbar sind. Diese Meinung spiegelt sich auch in dem aktuellen Trend wider, die DC-Methode bevorzugt in Marktforschung-, Umwelt-, Gesundheit- und Kulturbewertungsstudien einzusetzen.

### 2.3.3 Vergleich von Contingent Valuation und den multiattributiven Ansätzen

Die vorgestellten Befragungsmethoden CV, CA und DC unterscheiden sich in dem gewonnenen Informationsgehalt, in ihrer Komplexität und in der Eignung, theoriekonsistente Schätzungen der WTP zu liefern (Bateman et al., 2002; Hanley et al., 2001). Die CV-Methode, die die WTP der Befragten durch offene oder geschlossene Frage(n) direkt erfasst, ermöglicht die theoriekonsistente monetäre Bewertung von ganzheitlichen Umweltveränderungen. Im Vergleich dazu liefern die Ergebnisse aus CA und DC zusätzlich Information über die Wichtigkeit und ggf. den monetären Wert einzelner Merkmale des Umweltgutes. CA und DC weisen somit gegenüber CV den Vorteil auf, multidimensionalen Änderungen in der Umweltqualität Rechnung zu tragen (Bennett & Blamey, 2001). Die Ergebnisse aus CA und DC können damit erweiterte Handlungsmöglichkeiten in der Politik bieten: Sie geben z. B. Anhaltspunkte für die gezielte Gestaltung von Umweltmaßnahmen, die von möglichst vielen Menschen gut geheißen werden; wenn ein Unternehmen irreparable Schäden an einem Umweltgut verursacht hat, können sie als Grundlage für die Entscheidung dienen, welche alternative Umweltgüter und –services und in welcher Höhe das Unternehmen bereitstellen soll, um die Verluste der Gesellschaft zu kompensieren.

Komplexe umweltpolitische Entscheidungen erfordern meistens Kosten-Nutzen-Analysen verschiedener Umweltszenarien und nicht die generelle Entscheidung für oder gegen eine Maßnahme. Da in CV immer nur eine Maßnahme bewertet wird, wären im Rahmen komplexer Entscheidungen mehrere Studien notwendig. Somit stellen in diesem Fall CA und DC, die mehrere Szenarien bzw. Beschreibungen des Umweltgutes enthalten, eine kostengünstige Alternative dar, den monetären Wert von Umweltgütern und –services zu erfassen (Bennett & Blamey, 2001; Hanley et al., 2001). Die Ableitung theoriekonsistenter WTP-Schätzungen aus der traditionellen CA-Methode ist jedoch eingeschränkt möglich (s. Hanley et al., 1998; Louviere, 2001).

Neben den o.g. Vorteilen von CA und DC wird in der Forschungsliteratur diskutiert, dass die multiattributiven Ansätze möglicherweise weniger anfällig für verschiedene in CV-Studien aufgedeckte Verzerrungen sind (Alpizar et al., 2003; Hanley et al., 2001; Louviere, Hensher & Swait, 2000). Auf die Anfälligkeit der multiattributiven Ansätze für die hypothetische und

die strategische Verzerrung sowie auf die relevante empirische Evidenz wird jeweils in Abschnitten 3.1 und 5.2.3 näher eingegangen.

Die Anwendung der multiattributiven Ansätze kann neben den o. g. Vorteilen auch Nachteile aufweisen (Hanley et al., 2001; R.T. Carson et al., 1999; R.T. Carson & Groves, 2007). Ein Nachteil im Vergleich zur CV-Methode ist in der erhöhten Komplexität der multiattributiven Ansätze zu sehen: Von den Personen werden mehrere Wahlen bzw. Bewertungen zu mehreren Objekten mit mehreren Merkmalen und Merkmalsausprägungen erfordert. Empirische Studien haben gezeigt, dass die Aufgabenkomplexität einen negativen Einfluss auf die Entscheidungen der Befragten haben kann (Adamowicz, Louviere & Swait, 1998; Bradley, 1988; Brocke, 2006; Mazotta & Opaluch, 1995; Swait & Adamowicz, 1996). Die Komplexität beeinflusst den Fehlerterm im MNL-Modell, so dass mehr Komplexität mit mehr Zufallsfehlern in den Wahlen der Befragten assoziiert ist und sich negativ auf die Güte der geschätzten Modelle auswirkt (Brocke, 2006; DeShazo & Fermo, 2002).

Infolge der gestiegenen Aufgabenkomplexität können die Personen weiterhin Heuristiken bzw. Daumenregeln wie bspw. lexikografische Präferenzen oder Minimum- und Maximum-Strategien anwenden, um die Entscheidungsaufgabe zu vereinfachen. (Foster & Mourato, 2002; Sælensminde, 2002, 2006; Tversky, 1972; Tversky & Kahneman, 1974). Diese Vereinfachungsstrategien können dazu führen, dass die Personen zufriedenstellende Optionen wählen anstatt sich theoriekonform zu verhalten. Ermüdungs- und Lerneffekte in den multiattributiven Ansätzen, die für gewöhnlich die Befragten mit einer Reihe von Aufgaben konfrontiert, sind ebenfalls möglich (Adamowicz et al., 1998).

Die wiederholte Abfrage von Wahlen und Urteilen kann nach Ansicht einiger Autoren strategisches Verhalten begünstigen (R.T. Carson et al., 1999; R.T. Carson & Groves, 2007), sowie Anker- oder Reihenfolge-Effekte verursachen (vgl. Ariely, Loewenstein & Prelec, 2003; Kjær, Bech, Gyrd-Hansen & Hart-Hansen, 2006; Mørkbak, Christensen & Gyrd-Hansen, 2010). Das Fehlen einer Status-Quo-Option kann die WTP-Schätzungen aus den multiattributiven Ansätzen ebenfalls beeinträchtigen (Alpizar et al., 2003). Das Vorliegen einer Status-Quo-Option kann jedoch zum s.g. Status-Quo-Bias führen (W. Samuelson & Zeckhauser, 1988). Damit ist die unerwartet häufige Wahl der Status-Quo-Option gemeint und den damit assoziierten positiven Nutzen von dieser (Boxall, Adamowicz & Moon, 2009).

Die oben beschriebenen Effekte und Verzerrungen können die Reliabilität der Ergebnisse oder auch die theoretischen Annahmen der multiattributiven Ansätze verletzen und somit die Validität der Schätzungen aus diesen Methoden beeinträchtigen. Die Validität der verschiedenen Befragungsmethoden ist Gegenstand von Abschnitt 2.5.

## 2.4 Die Schätzung von Zahlungsbereitschaften aus den Befragungsmethoden

Die WTP-Schätzung aus den offenen Formaten in CV kann direkt aus der Verteilung der Daten anhand Mittelwerte oder Mediane erfolgen (Damschroder, Ubel, Riis & D.M. Smith, 2007; OECD, 2002). Die geschätzten Präferenzmodelle in CE, DC und CA können hingegen für unterschiedliche Zwecke eingesetzt werden (vgl. Lancsar, Louviere & Flynn, 2007 für einen Überblick). In der monetären Umweltbewertung sind dabei die Ermittlung von marginalen Substitutionsraten (MSR) bzw. marginalen WTP sowie die Bestimmung von Zahlungsbereitschaften für ganzheitliche Alternativen von besonderer Bedeutung.

Die MSR ist ein Maß dafür, wie viele Einheiten eines Merkmals nötig sind, um ein anderes Merkmal zu kompensieren. Die Ergebnisse von CA und DC lassen bspw. für die Befragten schätzen, wie viel Einschränkung bei der Zugänglichkeit eines Waldgebiets ( $X_2$ , in Tagen) akzeptabel wäre, um eine bedrohte Tierart zu schützen ( $X_1$ ). In Anlehnung an die Konsumententheorie errechnet sich die MSR als die partielle Ableitung der Nutzenfunktion in Bezug auf das erste und das zweite Merkmal und die Berechnung von ihrem Verhältnis (vgl. Hanley et al, 2001).

Die Berechnung marginaler Substitutionsraten ist in Gleichung 10 dargestellt, wobei  $V$  die geschätzte Nutzenfunktion ist,  $X_1$  und  $X_2$  Merkmale des Gutes oder Services sind und  $\partial$  die partielle Ableitung:

$$MSR(X_1, X_2) = (\partial V / \partial X_1) / (\partial V / \partial X_2) \quad (10)$$

Wenn die Nutzenfunktion linear additiv ist, entspricht die MSR in Gleichung 10 dem Verhältnis der geschätzten Präferenzparameter für  $X_1$  und  $X_2$  (Rosen & Small, 1981). Für die Ermittlung von MSR aus nicht linearen Nutzenfunktionen sind komplexere Berechnungen notwendig (vgl. Lancsar et al., 2007).

Wenn es sich bei einer der Variablen im Modell um den Preis oder um die Kosten handelt, so ist das Ergebnis mit dem Preis im Zähler von Gleichung 10 als der marginale (negative) Nutzen des Preises zu interpretieren und wird als „marginaler“ bzw. „impliziter“ Preis bezeichnet (Hanley et al., 2001). Damit lassen sich in den multiattributiven Ansätzen die einzelnen Merkmale eines Umweltgutes oder einer Schutzmaßnahme monetär bewerten.

Die Berechnung von WTP aus DC-Modellen wurde im Allgemeinen von Rosen und Small (1981) beschrieben. Die Berechnung von theoriekonsistenten WTP für ganzheitliche Alternativen aus der metrischen Conjoint Analyse ist hingegen mit Schwierigkeiten verbunden (Hanley et al., 1998, 2001). An dieser Stelle wird daher nur auf die Bestimmung der WTP für ganzheitliche Alternativen aus Wahlmodellen eingegangen.

Dabei lassen sich zwei verschiedene WTP-Maße gemäß zwei verschiedener Modelle bestimmen (Ryan, 2004). Die s. g. Modelle mit multiplen Alternativen betrachten mehrere verschiedene Auswahlalternativen und tragen Rechnung der Unsicherheit im Wahlmodell über die Wahl einer Option. Die WTP bezieht sich auf den Erwartungswert der multiplen Alternativen und ist für das MNL-Modell gegeben durch:

$$WTP = -\frac{1}{\beta_{Preis}} [\ln \sum_{j=1}^J \exp(V_j^0) - \ln \sum_{j=1}^J \exp(V_j^1)] \quad (11)$$

wobei  $J$  die Anzahl von Optionen im Wahlset darstellt, und  $V_j^0$  und  $V_j^1$  die Werte der geschätzten Nutzenfunktion für jede Option  $j$  entsprechend vor und nach der Veränderung (Lancsar & Savage, 2004).

In den s.g. „State-of-the-World“-Modellen werden die Nutzenniveaus einer einzigen Alternative der Status-Quo-Option gegenübergestellt. Das WTP-Maß errechnet sich dabei als:

$$WTP = -(V^0 - V^1) / \beta_{Preis} \quad (12)$$

Im „State-of-the-World“-Modell wird dabei angenommen, dass in der Tat nur eine Alternative realisiert werden kann, und die Person zwischen dem Status-Quo und dieser einen Alternative zu wählen hat. Die Werte der Merkmale im Status-Quo-Zustand und in dem verbesserten Umweltzustand werden dabei jeweils mit den geschätzten Teilnutzenwerten gewichtet und zusammenaddiert. Wenn die Berechnungen für verschiedene Merkmalsausprägungen bzw. verschiedenen Werte der Merkmale durchgeführt werden, lassen sich verschiedene Umweltveränderungen monetär bewerten und vergleichen. Gleichung 12 kann als ein Spezialfall von Gleichung 11 betrachtet werden, wenn die Wahlwahrscheinlichkeit für die Alternative 100% ist (vgl. Ryan, 2004).

Durch die Division der Koeffizienten bei den marginalen Substitutionsraten, die allen WTP-Maßen zugrunde liegen, spielen unterschiedliche Skalenparameter in verschiedenen Modellen keine Rolle. Die WTP-Schätzungen (marginal und für ganzheitliche Alternativen) aus verschiedenen Modellen lassen sich somit direkt miteinander vergleichen (Train, 2003).

## 2.5 Ansätze zur Überprüfung der Validität der Befragungsmethoden

Die Validität einer Methode betrifft die Frage, ob die Methode das Konstrukt erfasst, das sie beansprucht zu erfassen (Freeman III, 2003). Im Bereich der Umweltbewertung handelt es sich dabei um die wahre WTP der Befragten für ein Umweltgut (vgl. Abschnitt 2.2). Die Validität einer Methode beinhaltet verschiedene Aspekte – inhaltliche Validität, Kriteriumsvalidität, Konstruktvalidität – und kann mit verschiedenen Tests überprüft werden (vgl. Bateman et al., 2002; Freeman III, 2003; Mitchell & R.T. Carson, 1989). An dieser Stelle werden lediglich die Ergebnisse aus Studien vorgestellt, die die Kriteriums- bzw. konvergente Validität der Befragungsmethoden zum Gegenstand haben. Die externe Kriteriumsvalidität bezieht sich auf

die Gegenüberstellung der Ergebnisse einer Methode mit objektiven empirischen Kriterien, die das untersuchte Konstrukt abbilden. Die interne Kriteriumsvalidität beruht auf Vergleiche der Ergebnisse einer Methode mit dem Ergebnis einer anderen Methode, die beansprucht dasselbe Konstrukt valide zu erfassen. Werden die Ergebnisse mehrerer Methoden verglichen, die beanspruchen dasselbe Konstrukt zu messen, handelt es sich dabei um die konvergente Validität.

### 2.5.1 Interne Kriteriumsvalidität und konvergente Validität

Tests der internen Kriteriumsvalidität stellen eine der populärsten Formen der Validitätsüberprüfung der Befragungsmethoden dar. Dabei werden entweder die aus verschiedenen Formaten einer Methodenklasse ermittelten Präferenzen bzw. die Zahlungsbereitschaften miteinander verglichen (z.B. das offene vs. geschlossene CV-Format, Ranking vs. Rating in CA) oder die Ergebnisse verschiedener Befragungsmethoden untereinander (z.B. CV vs. CA, CA vs. DC). Diesen Tests liegt die Annahme zugrunde, dass die Schätzungen des monetären Wertes für ein Gut aus verschiedenen Frageformaten bzw. Befragungsmethoden übereinstimmen würden, wenn die Methoden die wahre WTP der Befragten valide erfassen würden.

Die Ergebnisse mehrerer empirischer Tests sprechen eher für den Gegenteil. Sowohl Studien aus der Umweltbewertung als auch Studien aus der Marktforschung, die die Befragungsmethoden miteinander vergleichen, zeigen bis auf wenige Ausnahmen (z.B. Boxall, Adamowicz, Swait, M. Williams & Louviere, 1996), dass die aus CA und DC geschätzten Zahlungsbereitschaften höher ausfallen als die Schätzungen der CV-Methode (Backhaus et al., 2005; Sattler & Nitschke, 2003; Stevens et al., 2000). In der Umweltbewertungsforschung z.B. übersteigen die Schätzungen aus CA und DC um das 0.6-fache (Magat, Viscusi & Huber, 1988) bis 5.0-fache (Stevens, Barrett & C.E. Willis, 1997) die Schätzungen aus CV.

Die WTP-Schätzungen aus verschiedenen Frageformaten bzw. Modi zum Abruf von Präferenzen innerhalb einer Methodenklasse stimmen ebenfalls selten überein. In CV bspw. fallen die Schätzungen aus dem geschlossenen Format im Durchschnitt doppelt so hoch aus wie die Schätzungen aus dem offenen Format; die Schätzungen des double-bounded-Formats fallen auch geringer aus als die Schätzungen aus dem geschlossenen Format (vgl. Abschnitt 2.3.1). Empirische Studien zur internen Validität von Ranking-, Rating- und wahlbasierten Conjoint-Aufgaben konnten ebenfalls keine konsistente Übereinstimmung der Zahlungsbereitschaften aufzeigen (z.B. Boyle, Holmes, Teisl & Roe, 2001; Caparrós, Oviedo, & Campos, 2008; Kallish & Nelson, 1991).

Für diesen s.g. *elicitation effect* existieren eine Reihe von möglichen Erklärungen (s. Stevens et al., 2000). Verzerrungen in den Antworten zu einzelnen oder mehreren Befragungsmethoden können z.B. auf das Fehlen von Anreizen für die wahrheitsgemäße Äußerung der Zahlungsbereitschaft zurückgeführt werden (R.T. Carson & Groves, 2007; R.T. Carson et al.,

1999; Ding et al., 2005) und an strategischem Verhalten liegen. Eine weitere Hypothese unterstellt, dass die Präferenzen und WTP der Befragten nicht stabil sind und die verschiedenen Methoden unterschiedliche Konstruktionsprozesse in Gang setzen (vgl. Gregory, Lichtenstein & Slovic, 1993). Unterschiede in den Schätzungen der Befragungsmethoden wurden ebenfalls auf Unterschiede in der Aufgabenkomplexität einzelner Methoden zurückgeführt (DeShazo & Fermo 2002; Hanemann, 1994) sowie auf verschiedene psychologische Effekte wie z.B. Ankereffekte, Ja-Sagen oder Framingeffekte in den verschiedenen Frageformaten (D. Green et al. 1998; Holmes & Kramer, 1995; Saelensminde, 2006).

Ungeachtet der Ursachen für den *elicitation effect* ist das auf internen Vergleichen basierende Vorgehen der Validitätsüberprüfung problematisch. Es erlaubt keine Aussagen darüber, ob und welche Methode die wahre WTP der Befragten am genauesten erfasst, da ein objektives Kriterium für die wahre WTP fehlt. Das Heranziehen objektiver, externer Kriterien soll dabei helfen, die Validität der Befragungsmethoden besser zu beurteilen.

### 2.5.2 Externe Kriteriumsvalidität

Eine weitere Möglichkeit, die Validität der Befragungsmethoden zu überprüfen, besteht in dem Vergleich ihrer Ergebnisse mit einem externen Kriterium wie z.B. beobachtbarem Verhalten in realen Kaufsituationen oder realen Zahlungen in Experimenten. In zahlreichen Studien im CV-Bereich wurden die geäußerten Zahlungsbereitschaften für Konsum- und seltener für Umweltgüter mit tatsächlich geleisteten Zahlungen in realen oder experimentellen Märkten verglichen (z.B. Champ et al., 1997; Cummings et al., 1995; Johannesson, Liljas & Johannesson, 1998; Seip & Strand, 1992). In den meisten Studienergebnissen, die ausführlich in Abschnitt 3.1 präsentiert werden, zeichnet sich die Tendenz ab, dass die WTP aus hypothetischen CV-Befragungen höher ausfällt als reale Zahlungen. Diese Inkonsistenz zwischen hypothetischem und realem Verhalten, die als hypothetische Verzerrung bekannt ist, wird als einen wichtigen Hinweis für die fehlende Validität der CV-Methode interpretiert (vgl. Abschnitt 3.3.1).

Für CA und DC liegen noch relativ wenige Studien vor, die die hypothetischen Antworten der Befragten mit realen Entscheidungen vergleichen; die Ergebnisse dieser Studien sind insgesamt gemischt (vgl. Abschnitt 3.1 für eine ausführliche Darstellung). Die neuesten Erkenntnisse aus der Forschungsliteratur weisen darauf hin, dass möglicherweise verschiedene Befragungsmethoden in einem unterschiedlichen Ausmaß von der hypothetischen Verzerrung betroffen sind (Murphy et al. 2005; Voelckner, 2006). Eine systematische Übersicht der Forschung zur hypothetischen Verzerrung wird in Kapitel 3 gegeben.

In den meisten Studien, die die externe Kriteriumsvalidität der Befragungsmethoden zum Gegenstand haben, wird angenommen, dass die realen Zahlungen als fehlerfreie Messungen der wahren WTP herangezogen werden können (Murphy et al., 2005). Die Validität von Verhal-

tensdaten aus Situationen mit finanziellen Anreizen wurde jedoch von verschiedenen Autoren in Frage gestellt. Sattler & Nitschke (2003) z.B. kritisieren die WTP-Schätzungen aus realen Marktdaten, da diese für gewöhnlich auf Reaktionen auf Preise basieren, die in engen Grenzen variieren. Solche Daten ermöglichen lediglich eine eingeschränkte Aussage, ob die WTP des Käufers bei dem Kaufpreis oder höher liegt und liefern eher eine Untergrenze der wahren WTP. In Laborexperimenten, in denen die Produkte und die Preise systematisch variiert werden, können die tatsächlich geleisteten Zahlungen ebenfalls ein verzerrtes Bild der wahren WTP geben (Nagle & Holden, 2002). In ökonomischen Experimenten wird Kaufverhalten dadurch induziert, dass die Probanden für die Teilnahme Geld bekommen, das sie für Produkte/Güter im Experiment ausgeben können. In diesem "künstlichen" Kontext, in dem die Probanden möglicherweise kein echtes Kaufinteresse für das Produkt haben und nicht ihr eigenes Geld ausgeben, können strategisches Verhalten oder Zocken („gambling“) hervorgerufen werden anstatt wahre Präferenzen (Pesheva, Kroll & Vogt, 2011).

Die mithilfe von theoretisch anreizkompatiblen Vickrey-Auktionen und BDM-Lotterien hervorgerufenen realen Zahlungsbereitschaften in Feld- und Laborstudien sind ebenfalls kritisch zu werten. Ungeachtet der erwarteten Angabe der wahren WTP in diesen Methoden, zeigen die Ergebnisse einschlägiger empirischer Studien, dass sie in der Tat zu Unter- und Überbieten führen können und keine übereinstimmenden Schätzungen erbringen (z. B. Kagel, 1995; Kagel & Levin, 1993; Noussair, Robin, & Ruffieux, 2004). Folglich können die WTP-Schätzungen aus solchen anreizkompatiblen Verfahren nicht ohne weiteres als objektives Kriterium für die wahre WTP der Befragten herangezogen werden.

Reale Zahlungen in Verbindung mit nicht anreizkompatiblen Frageformaten können weiterhin strategisches Verhalten in der realen Situation und somit verzerrte WTP-Angaben begünstigen (R.T. Carson et al., 1999; R.T. Carson & Groves, 2007). Das Finden von einem tatsächlich objektiven Kriterium für die wahre WTP der Befragten gestaltet sich somit schwierig. Eine mögliche Alternative, die Validität von Methoden zur Messung von WTP zu testen, besteht darin, ein objektives Kriterium für die wahre WTP der Befragten experimentell zu erschaffen.

### 2.5.3 Experimente mit induzierten Präferenzen

Für Validitätsstudien im Rahmen der WTP- bzw. Präferenzmessung ist die Kontrolle der Vorlieben von den Personen von essentieller Bedeutung. Dies kann erreicht werden, indem die natürlichen Präferenzen der Personen „neutralisiert“ und an ihrer Stelle bekannte Präferenzen induziert werden. Dazu können Experimente mit induzierten Werten herangezogen werden (V.L. Smith, 1976). Die Idee von diesen Experimenten ist einfach: Den Probanden werden ihre wahren Werte im Experiment vom Forscher vorgegeben und sind diesem somit bekannt. Die vorgegebenen wahren Werte können (bei erfolgreicher Induktion) als Kriterium für die

Genauigkeit der Befragungsmethoden unter verschiedenen Bedingungen dienen. In diesem Abschnitt werden zwei alternative Ansätze zur Induktion von Präferenzen und WTP vorgestellt und diskutiert - das traditionelle Induzierte-Werte-Experiment aus der experimentellen Ökonomie (V.L. Smith, 1976) und das s. g. Principal-Agent-Paradigma aus der Entscheidungsforschung (Huber et al., 2002).

#### *Methode der induzierten Werte von V.L. Smith*

Die s.g. Induzierte-Werte-Experimente gehen auf die Arbeit von V.L. Smith (1976) zurück. Als Mittel zur Induktion von Präferenzen dient Belohnung, die drei Kriterien erfüllen muss, damit eine zuverlässige Induktion gewährleistet ist: *Monotonie* erfordert, dass man stets mehr Belohnung vor weniger Belohnung vorzieht und nie „gesättigt“ wird; *Salienz* erfordert, dass die Höhe der Belohnung direkt von dem Verhalten des Probanden abhängt; *Dominanz* setzt voraus, dass die Höhe der Belohnung ausreicht, um alleine das Verhalten der Probanden im Experiment zu beeinflussen und andere Motive eher unwahrscheinlich zu machen. In ökonomischen Experimenten wird Geld als Belohnungsmittel herangezogen, da es in hohem Maße die o.g. Kriterien erfüllt.

In einem typischen Induzierte-Werte-Experiment bekommen die Teilnehmer von dem Studienleiter einen Wert, den s.g. induzierten Wert, für ein fiktives Gut zugewiesen. Dies kann z. B. in Form von Tokens, Spielgeld oder Murmeln erfolgen, die beim Studienleiter am Ende des Experiments gegen Geld getauscht werden. Ausgehend von ihrem persönlichen induzierten Wert für das fiktive Gut, hat die Person bspw. in einer Auktion für das Gut zu bieten oder zu entscheiden, ob sie das Gut zu einem Preis erwerben möchte. Wenn ein Proband das Gut erwirbt, dann bekommt er die Differenz zwischen seinem induzierten Wert und dem Kaufpreis ausgezahlt.

Das Prinzip der Induzierte-Werte-Experimente kann im CV-Kontext am Beispiel der Studie von L.O. Taylor et al. (2001) veranschaulicht werden. Die Studie hatte das Ziel die Genauigkeit von CE in einer Bedingung mit realen Zahlungen und in einer hypothetischen Bedingung zu testen. Für die Teilnahme am Experiment bekam jeder Proband 7 \$ Startkapital und die Chance weiteres Geld zu verdienen. Die Höhe des verdienten Betrages hing von der Entscheidung ab, ob ein nicht näher beschriebenes fiktives Umweltprojekt zum Preis von 5 \$ realisiert wird oder nicht. Den Probanden wurden zuerst induzierte Werte zwischen 1 \$ und 10 \$ verteilt, die dem persönlichen Nutzen eines jeden Probanden durch die Realisierung des Projektes entsprachen. Die Probanden hatten dann für oder gegen die Realisierung des fiktiven Projektes in einem Referendum abzustimmen. Wenn sich die Mehrheit für das Projekt entschied, hatte jeder in der Gruppe 5 \$ zu bezahlen, unabhängig von seinem induzierten Wert. In diesem Fall wurden am Ende des Experiments jeder Person 7 \$ Startkapital zzgl. der Differenz zwischen ihrem induzierten Wert und den 5 \$ Kosten für die Realisierung des Pro-

jekt es ausgezahlt. Wenn sich die Mehrheit gegen das fiktive Projekt entschied, wurde dieses nicht realisiert und alle Probanden bekamen nur die 7 \$ Startkapital ausgezahlt. Rationale nutzenmaximierende Entscheider mit induzierten Werten über 5 \$ und erwarteten Gewinnen sollten demnach im Referendum stets mit „Ja“ abstimmen. Personen mit induziertem Wert kleiner 5 \$ sollten hingegen stets mit „Nein“ abstimmen, weil die Bereitstellung für sie mit Verlusten verbunden wäre. Da die wahren Werte der Probanden bekannt sind, erlaubt diese Methode somit direkte Tests, ob sich eine Person rational, also gemäß dem induzierten Nutzen, verhält oder nicht.

Im CV-Kontext ist die Bewertung für ganzheitliche Güter von Interesse, daher ist der Ansatz von V.L. Smith leicht anzuwenden, auch wenn das „Gut“ sehr abstrakt ist. Durch die Vorgabe induzierter Werte besitzen die Personen ihnen bekannte und mit keinerlei Unsicherheit verbundene Präferenzen für das Gut. Relevant für die Entscheidung der Probanden ist somit alleine die Differenz zwischen dem individuellen induzierten Wert und den Kosten für die Bereitstellung des fiktiven Gutes. Die Entscheidungsaufgabe wird somit im CV-Kontext auf eine einfache Rechenaufgabe reduziert. In DC und CA werden die Güter jedoch mithilfe von mehreren Merkmalen und variierenden Merkmalsausprägungen definiert, daher sind induzierte Werte für jede Merkmalsausprägung zu spezifizieren und dem Probanden zu vermitteln. Die Anwendung des typischen Induzierte-Werte-Experiments im Rahmen der multiattributiven Befragungsmethoden ist daher im Vergleich zu CV um einiges komplexer und relativ selten.

Ein Beispiel für die Anwendung der Methode im DC-Kontext liefert die Studie von Collins und Vossler (2009). Diese untersuchen die Konsistenz von DC-Wahlen und die Güte der WTP-Schätzungen in Abhängigkeit von verschiedenen Bereitstellungsregeln und von der Anzahl der Optionen im Choice-Set (eine vs. zwei und Status-Quo). Das „Gut“ im Experiment wurde mithilfe von 3 Attributen (Rot, Blau und Preis) mit jeweils 2 Ausprägungen beschrieben. Das Merkmal „Rot“ hatte die Ausprägungen 1 und 2, das Merkmal „Blau“ die Ausprägungen 1 und 3, der Preis für die Optionen war 2.50 \$ oder 4.00 \$. Die Status-Quo-Option besaß auf den Merkmalen „Rot“ und „Blau“ jeweils die Ausprägung 1 und auf die Preisvariable stets die Ausprägung 0 \$.

Die persönlichen Werte der Studienteilnehmer variierten zufällig je Merkmal „Blau“ und „Rot“ zwischen 1 \$, 2 \$ und 3 \$ je Merkmalsausprägung. Es wurde eine additive Nutzenfunktion angenommen, so dass sich der Gesamtnutzen einer Option für eine Person aus der Summe der individuellen Teilnutzenwerte für die Merkmalsausprägungen ergab. Eine Person mit einem induzierten Wert für Rot = 2 \$ und Blau = 3 \$ würde sich bei der Wahl zwischen einer Alternative mit den Merkmalsausprägungen Rot = 2, Blau = 2 und Preis = 2.5 \$ und der SQ-Option (Ausprägungen: Rot = 1, Blau = 1, Preis = 0 \$), für die Alternative und gegen SQ entscheiden (Alternative:  $2 \cdot 2\$ + 2 \cdot 3\$ - 2.5\$ = 7.5\$$  vs. SQ:  $1 \cdot 2\$ + 1 \cdot 3\$ - 0\$ = 5\$$ ).

Wie dieses Beispiel zeigt, besteht die Aufgabe der Probanden im Kontext von multiattributiven induzierten Werten darin, den Gesamtwert der präsentierten Alternativen im Choice-Set auszurechnen und die Alternative mit dem höchsten positiven Nutzen zu wählen. Bevor die Studienteilnehmer die tatsächlichen Aufgaben in der Studie von Collins & Vossler (2009) bearbeiteten, übten sie daher zuerst in einem beispielhaften Choice-Set die Gewinne aus einzelnen Optionen zu berechnen.

Die Ergebnisse beider vorgestellten Studien und weiterer Experimente mit induzierten Werten werden in Abschnitt 3.2 detailliert dargestellt und diskutiert. An dieser Stelle ist nur das Design der Experimente von Interesse, um Vor- und Nachteile der vorgestellten Induktionsmethode aufzeigen zu können.

Wie die o.g. Beispiele im CV- und DC-Kontext deutlich machen, weisen Induzierte-Werte-Experimente verschiedene Einschränkungen auf: Zum einen erfolgt die Induktion von wahren Werten für ein nicht näher beschriebenes oder sehr abstraktes Gut. Zum anderen entspricht das abstrakte aber zugleich sehr vereinfachte Experimentaldesign im Kontext der multiattributiven Methoden nicht der tatsächlichen Komplexität realer DC- und CA-Befragungen, die für gewöhnlich Güter mit mehreren Merkmalen und mehreren Merkmalsausprägungen zum Gegenstand haben. Weiterhin handeln die Teilnehmer in Induzierte-Werte-Experimenten in vollkommener Sicherheit, da sie ihre wahren Werte genau kennen und ihre Gewinne exakt berechnen können. Die Studienteilnehmer werden bei komplexeren Aufgaben wie DC sogar im Vorfeld in der „Berechnung“ von Werten trainiert; im Fall von den CV-Formaten werden die Entscheidungen der Probanden auf einfache Rechenaufgaben reduziert.

Das Induktionsvorgehen von V.L. Smith ist in Übereinstimmung mit der theoretischen Annahme, dass Personen über ihre Präferenzen genauestens Bescheid wissen und mit Sicherheit den Preis benennen können, den sie bereit sind maximal für eine Einheit des Gutes zu bezahlen (Hanemann, 1984). In Ökonomie und Psychologie wird jedoch aktuell von mehreren Forschern die Ansicht vertreten, dass die Befragten an der Stelle eines wahren WTP-Wertes eine Verteilung von möglichen Werten besitzen (s. T. Wang, Venkatesh & Chatterjee, 2007 für eine Zusammenfassung dieser Ansätze). Bei der Bewertung von (Umwelt-) Gütern, die trade-offs zwischen bspw. Umweltmerkmalen und Kosten abverlangen, empfinden die Personen daher mehr oder weniger Unsicherheit über den wahren Wert des Gutes (Gregory, Lichtenstein, T.C. Brown, Peterson & Slovic, 1995; Gregory & Slovic, 1997; Li & Mattsson, 1995; Opaluch & Segerson, 1989; H. Wang, 1997). In der Induktionsmethode von V.L. Smith wird dieser möglichen Präferenzunsicherheit in natürlichen Bedingungen keine Rechnung getragen. Die Ergebnisse aus Induzierte-Werte-Experimenten weisen somit möglicherweise eine eingeschränkte prädiktive Validität für Verhalten außerhalb des Labors auf.

*Induktion von Präferenzen und Zahlungsbereitschaften mithilfe des Principal-Agent-Paradigmas*

Das Principal-Agent-Paradigma (PAP) ist ein Forschungsansatz zur Induktion von Präferenzen, der im Rahmen der Entscheidungspsychologie (Ariely, 2000; Huber et al., 2002; N.M. Klein & Bither, 1987) und in der Präferenzforschung (Reiners, 1996; Teichert, 2000) Anwendung findet. Die Induktion von Präferenzen in PAP bezieht sich auf multiattributive Objekte. Dabei wird allen Probanden die Präferenzstruktur einer fiktiven Person (*Principal*) vermittelt. Die Probanden erlernen in Trainings oder erschließen aus Grafiken, welche Merkmale und Merkmalsausprägung dem Principal mehr wert sind und um wie viel mehr wert sind als andere Merkmale bzw. Ausprägungen. Anschließend haben die Probanden verschiedene Präferenzurteile entsprechend der Präferenzen des Principals bzw. als seine *Agenten* zu fällen. Die vorgegebenen Präferenzen der fiktiven Person, die als Referenz für die Entscheidungen und Urteile der Probanden im Experiment dienen, werden als die wahren Präferenzen betrachtet. Somit lassen sich in PAP ähnlich dem Induktionsansatz von V.L. Smith systematische Unterschiede zwischen den „wahren“ vorgegebenen und unter verschiedenen Bedingungen geäußerten Präferenzen identifizieren und quantifizieren (z.B. Huber et al., 2002; Teichert, 2000).

Die Anwendung von PAP setzt die zuverlässige Vermittlung der fiktiven Präferenzstruktur voraus. Diese ist unter anderem durch die Anwendung verschiedener Trainings zur Induktion fremder Präferenzen möglich (Großmann, Holling, Brocke, Graßhoff & Schwabe, 2005; Großmann et al., 2007; Huber & Elrod, 1981; Meyer, 1987; P.M. West, 1996). Diese Trainingsprozeduren unterscheiden sich sowohl in dem Ausmaß, in dem sie den natürlichen Prozess des Erwerbs von Präferenzen widerspiegeln als auch in dem Erfolg, mit dem fremde Präferenzen induziert werden.

Ein von Großmann et al. (2005) entwickeltes Trainingsverfahren zur Induktion fremder Präferenzen konnte im Vergleich zu älteren Studien zuverlässige Induktionserfolge erzielen (vgl. Huber & Elrod, 1981; Meyer, 1987; P.M. West, 1996). Das Training wurde positiv in verschiedenen empirischen Tests evaluiert – die in dem Training erworbenen Präferenzen sagten in einem sehr hohen Ausmaß die induzierten wahren Präferenzen vorher ( $R^2 = .97$  und  $R^2 = .99$ ; vgl. Großmann et al., 2005; Großmann et al., 2007). Im Vergleich zu früheren Induktionsstudien weist das Training ebenfalls den Vorteil auf, den Erwerbsprozess der fremden Präferenzen ähnlich der natürlichen Entwicklung von Präferenzen zu gestalten. Die Probanden lernen dabei in einem induktiven Lernprozess die Präferenzen der fiktiven Person beherrschen (vgl. für eine ähnliche Auffassung in der ökonomischen Forschung die „Preference Discovery“-Hypothese von Plott, 1996). Das Training von Großmann und Kollegen, das eine präzise und dem natürlichen Erwerbsprozess von Präferenzen ähnliche Vermittlung der frem-

den Präferenzstruktur gewährleistet, findet Anwendung in der vorliegenden Arbeit und wird näher in Abschnitt 3.5 präsentiert.

Im Rahmen von PAP werden bislang lediglich Präferenzen bzw. Nutzenwerte für Optionen induziert, aber keine Zahlungsbereitschaften. Wenn die Probanden bspw. nach dem Training präzise entscheiden können, welche Option dem Principal mehr Nutzen stiftet, so geht es im Fall der Zahlungsbereitschaft auch darum, diesen Unterschied monetär ausdrücken zu können. Für welche von zwei Optionen würde sich bspw. ein Principal entscheiden, wenn er eine hochwertige Option A zur Auswahl hat (Gesamtnutzenwert von 10) zum Preis von 50 € und eine eher geringwertige Option B (Gesamtnutzenwert von 2) zum Preis von 5 €?

Wie in Abschnitten 2.1 beschrieben wurde, steht das Nutzenkonzept in einem direkten Zusammenhang mit dem WTP-Konzept. Die WTP korrespondiert direkt mit dem wahrgenommenen Nutzen und stellt dessen monetäre Äquivalent (Balderjahn, 2003) bzw. in monetären Einheiten ausgedrückte Präferenzen dar (McFadden & Leonard, 1993). Varian (2003) begründet theoretisch den proportionalen Zusammenhang zwischen dem Nutzen von einem Gut  $X$  ( $U_x$ ) und der Zahlungsbereitschaft für das Gut ( $P_x$ ):

$$k = U_x/P_x \quad (13)$$

Der Faktor  $k$  präsentiert dabei die Austauschrate zwischen Nutzen und Geld. Gleichung 13 ist dabei lediglich eine andere Darstellung von Gleichung 10. Ist z.B.  $k = 0.1$ , so bedeutet das, dass ein Anstieg des Nutzens um 1 Einheit mit einem Anstieg der WTP um 10 Einheiten assoziiert wird. Wenn  $k$  und die Nutzenwerte der Optionen bekannt sind, kann die theoretisch wahre WTP für jede Option berechnet werden.

Um die Frage zu beantworten, welche Option ein Principal in PAP wählen würde (Option A: Gesamtnutzenwert von 10 und Preis 50 € oder Option B: Gesamtnutzenwert von 2 und Preis von 5 €), ist es essentiell die Austauschrate zwischen Nutzen und Geld zu kennen. Bei  $k = 0.1$  hätte die Option A gemäß Formel 13 eine wahre WTP von 100 € ( $10/0.1$ ), und die Option B hätte den wahren Wert von 20 € ( $2/0.1$ ). Die Option A würde dem Principal 50 € Nutzen stiften (100 € wahre WTP – 50 € Kosten), während Option B nur 15 € Nutzen stiften würde (20 € wahre WTP – 5 € Kosten). Somit würde ein Principal mit  $k = 0.1$  Option A wählen. Bei  $k = 0.25$  würde ein Principal hingegen Option B wählen. Ein Principal bzw. seine Agenten könnten sich in PAP somit unterschiedlich entscheiden in Abhängigkeit von  $k$ .

Um wahre WTP in PAP zu induzieren, sind somit zwei Voraussetzungen zu erfüllen: Der Proband hat den Nutzen zu kennen, den verschiedene Optionen für den Principal besitzen. Diese Voraussetzung wird bspw. durch das Training zur Induktion fremder Präferenzen von Großmann et al. (2005) gewährleistet. Zum anderen haben die Probanden die Austauschrate zwischen Nutzen und Geld vom Principal zu kennen.

Die Austauschrate des Principals könnte den Probanden direkt vorgegeben werden. Dabei würde sich die Aufgabe der Probanden jedoch ähnlich dem Ansatz von V.L. Smith darauf reduziert, aus den gelernten Präferenzen die WTP gemäß Formel 13 auszurechnen. Studien in der Konsumentenforschung zeigen jedoch, dass die Personen ihre Präferenzen in monetären Einheiten ausdrücken, indem sie nach einem Punkt auf der monetären Skala suchen, der ihrem Nutzen von dem Produkt entspricht (Chernev, 2003; Janiszewski & Lichtenstein, 1999; Lynch, Chakravarti & Mitra, 1991). Dies geschieht durch Angleichung der Endpunkte der Nutzen- und der monetären Skala und durch Anpassung der Änderung in der Nutzenskala durch proportionale Änderung auf der monetären Skala (vgl. Chernev, 2003). Um diesen Prozess der WTP-Findung möglichst genau in der vorliegenden Arbeit abzubilden, wird auf die direkte Vorgabe von  $k$  verzichtet. Stattdessen wird in dieser Arbeit die Austauschrate zwischen Nutzen und Geld für die fiktive Person *indirekt* vermittelt, indem den Probanden lediglich die wahre WTP für die beste und für die schlechteste Option als Ankerwerte genannt werden.

Die Eignung des o.g. Vorgehens zur Induktion von wahren Zahlungsbereitschaften wurde im Rahmen von zwei empirischen Studien bestätigt (Todorova, 2004). Die Probanden in diesen Studien konnten auf Basis der erlernten Präferenzen und der vorgegebenen wahren Referenzwerte, die wahre Zahlungsbereitschaft der fiktiven Person für weitere Optionen sehr genau ableiten.

#### *Vergleich von beiden Ansätzen*

Die beschriebenen Ansätze zur Induktion von Präferenzen (Induzierte-Werte-Experimente und PAP) haben den Vorteil gemein, experimentell die natürlichen Präferenzen der Probanden zu neutralisieren und an ihrer Stelle dem Forscher bekannte und von diesem kontrollierte Präferenzen zu induzieren. Beide Ansätze unterscheiden sich jedoch grundsätzlich voneinander darin, wie die Induktion von Präferenzen erfolgt.

Während beim Ansatz von V.L. Smith es sich um individuelle Präferenzen für ein fiktives abstraktes Gut handelt, werden im PAP homogene Präferenzen für konkrete multiattributive Güter induziert. Insofern eignet sich PAP besonders gut für Forschungszwecke im Rahmen multiattributiver Verfahren. Im Ansatz von Smith werden den Probanden die wahren Werte für das Gut bzw. für seine Merkmale direkt genannt, so dass sich die Aufgabe der Probanden im Experiment auf eine einfache Kosten-Nutzen-Rechnung reduziert. In PAP werden die Präferenzen der Befragten hingegen als das Ergebnis von einem induktiven Lernprozess aufgefasst. In Übereinstimmung mit der Annahme, dass Präferenzen das Ergebnis von Erfahrung sind, werden im PAP die Präferenzen der fiktiven Person in Trainings angeleert.

Weiterhin unterscheiden sich beide Ansätze darin, in welche Form sie Präferenzen induzieren. Im Ansatz von Smith werden die Präferenzen in monetären Einheiten vorgegeben. Die indu-

zierten Nutzenwerte in PAP sind hingegen nicht monetärer Natur – die wahre WTP ist auf Basis der verinnerlichten Präferenzen und der Vorgabe von zwei Ankerwerten abzuleiten. Beim Erwerb der Präferenzen und bei der Ableitung der wahren WTP besteht somit im Rahmen von PAP eine größere Realitätsnähe. Im Vergleich zu einem Induzierte-Werte-Experiment, wo sich die Nutzenwerte der Probanden genau berechnen lassen und die Präferenzen mit keinerlei Unsicherheit verbunden sind, entscheiden die Probanden in einem PAP-Experiment nicht unter vollkommener Sicherheit. Vor diesem Hintergrund erscheint PAP eine interessante und auf den neuesten Erkenntnissen der Präferenzforschung basierende alternative Methode zur Induktion von Präferenzen und von wahren WTP zu sein.

## **2.6 Ziele der Arbeit**

In der vorliegenden Arbeit wird das oben beschriebene PAP in drei Studien verwendet, um die Validität der in Abschnitt 2.3 präsentierten Befragungsmethoden unter verschiedenen experimentalen Bedingungen zu testen. Studie 1 hat das Ziel, den Einfluss von einem realen finanziellen Anreiz vs. einem hypothetischen Anreiz auf die Güte von CV, CA und DC zu untersuchen sowie die Anfälligkeit der Methoden für die s. g. hypothetische Verzerrung zu vergleichen. In Studie 2 wird der Frage nachgegangen, welchen Einfluss verschiedene Personenmerkmale direkt und in Interaktion mit einem hypothetischen bzw. realen finanziellen Anreiz auf die Genauigkeit der WTP-Schätzungen aus den o. g. Befragungsmethoden haben. Studie 3 hat zum Gegenstand, die Anfälligkeit der Befragungsmethoden für strategische Antworten zu bestimmen und zu vergleichen. Die drei Studien werden getrennt in den folgenden drei Kapiteln präsentiert. Eine abschließende Zusammenfassung und Diskussion der Ergebnisse aller drei Studien bietet Kapitel 6 an.

Die Güte verschiedener Befragungsmethoden zur  
Umweltbewertung  
Ein Vergleich der Validität von Contingent Valuation,  
Conjoint- und Discrete Choice-Analyse  
Todorova, M.  
2016, XX, 304 S. 17 Abb., Softcover  
ISBN: 978-3-658-13517-1