Response Anomalies in Discrete Choice Experiments for Environmental Valuation

The Influence of Task Complexity, Attributes Thresholds and Survey Consequentiality

vorgelegt von

Dipl. rer. pol. techn.

Malte Oehlmann

von der Fakultät VI - Planen Bauen Umwelt

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften

- Dr. rer. oec. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Stefan Heiland

Gutachter: Prof. Dr. Volkmar Hartje

Gutachterin: Prof. Dr. Katrin Rehdanz

Tag der wissenschaftlichen Aussprache: 05. April 2019

Berlin 2019

Abstract

The use of discrete choice experiments (DCEs) to value environmental goods and services has been rapidly increasing over the last years. However, many studies call into question the validity of this approach. Biases in welfare estimates derived from DCEs may arise from response behavior which is not fully consistent with one or more of the assumptions underlying stated preference approaches. Since preferences are elicited in surveys, the central question is whether anomalous response behavior is triggered by the survey design. This cumulative dissertation analyzes response anomalies in DCEs for environmental valuation. By designing and evaluating two large-scale, nation-wide online surveys each time implementing a split sample approach, the influence of choice task complexity, attribute thresholds and survey consequentiality is investigated systematically.

First, the impact of choice task complexity on survey drop-outs, choice consistency, the attendance to attributes and the status quo effect is studied by systematically varying the design dimensions of the DCE following a design plan originally introduced in transportation research. It is shown that the complexity of the DCE impacts marginal as well as non-marginal welfare estimates. With respect to each individual design dimension, survey drop-outs are found to be positively related to the number of choice sets, alternatives and attributes. Choices tend to become more consistent the more choice sets are presented as well as with a narrow level range. While the attendance to attributes is negatively affected by the number of choice sets and alternatives, status quo choices are observed to decrease with the number of alternatives and increase with the number of choice sets, the level range and the similarity between alternatives.

Second, stated information on attribute cut-offs is used to study threshold-based decision making. Results suggest that prior cut-off elicitation influences willingness to pay estimates. Moreover, respondents are observed to employ cut-offs, which they are, however, in many cases willing to violate when having to make trade-offs in the DCE.

Third, it is investigated whether information on how survey results would be used as well as perceived survey consequentiality affect preferences. It is observed that neither the information about, nor the perception of survey consequentiality impact preferences.
Overall, the results of this dissertation suggest the design of the survey, in general, and the complexity of the DCE, in particular, might influence preferences. This emphasizes the need to further identify and accommodate heterogeneous decision making in DCEs and discrete choice models.

Zusammenfassung

Choice Experimente (CEs) wurden in den letzten Jahren verstärkt zur Bewertung von Umweltgütern und –dienstleistungen herangezogen. Viele Studien stellen jedoch die Validität dieses Ansatzes infrage. Insbesondere Antwortvorhalten, welches eine oder mehrere Annahmen geäußerter Präferenzerfassung verletzt, kann die Ergebnisse beeinflussen. Da Präferenzen aus Umfragen abgeleitet werden stellt sich die Frage, inwieweit das Design der Umfrage selbst solche Antwortanomalien hervorrufen kann. Diese kumulative Dissertation untersucht Antwortanomalien in CEs zur Umweltbewertung. Unter Zuhilfenahme zweier bundesweiter Bevölkerungsumfragen wird der Einfluss von Komplexität, Attribut-Cutoffs und der Verwendung der Umfrageergebnisse untersucht.

Zunächst wird gezeigt, wie sich die Komplexität des CEs, die einem Ansatz aus der Verkehrsökonomie folgend anhand der Designdimensionen systematisch variiert wird, auf die Abbruchrate, die Konsistenz der Auswahlentscheidungen, die Berücksichtigung von Attributen und den Status Quo-Effekt auswirkt. Hierbei zeigt sich, dass die Komplexität des CEs die ermittelten Wohlfahrtsmaße signifikant beeinflusst. Mit Bezug auf die einzelnen Designdimensionen ist zu beobachten, dass die Abbruchrate mit der Anzahl an Auswahlsets, Alternativen und Attributen zunimmt. Die Auswahlkonsistenz steigt mit der Anzahl der Auswahlsets und mit einem geringeren Spannweite zwischen den Attributsleveln, während die Anzahl berücksichtigter Attribute mit der Anzahl der Alternativen und Auswahlsets abnimmt. Darüber hinaus neigen die Befragten eher dazu die Status Quo-Alternative zu wählen, je mehr Auswahlsets präsentiert werden und je größer die Levelspannweite ist.

In einem zweiten Schritt wird der Einfluss von Attribut-Cutoofs untersucht. Die Ergebnisse zeigen, dass die Abfrage von Cutoffs vor dem CE zu höheren Zahlungsbereitschaften führt. Zudem nutzt eine Vielzahl der Befragten Cutoffs zur Entscheidungsfindung, wobei die Befragten in vielen Fällen dazu bereit sind, diese im CE zu verletzen.

Abschließend wird untersucht, ob Informationen über die Verwendung der Umfrageergebnisse sowie die wahrgenommenen Folgen der Antworten Einfluss auf Präferenzen haben. Für beides konnte kein signifikanter Effekt festgestellt werden.

Zusammenfassend zeigt die vorliegende Arbeit, dass das Umfragedesign im Allgemeinen und die Komplexität des CE im Speziellen Präferenzen beeinflussen können. Dieser Befund betont die Notwendigkeit, die Nutzung verschiedener Entscheidungsstrategien in CE und in der Modellierung diskreter Auswahlentscheidungen verstärkt in den Blick zu nehmen.

Table of Contents

1 Introduction

Climate change, air pollution or the loss of biodiversity are only some of the major challenges our society has to face. Policy makers are expected to adopt a variety of environmental measures to approach these problems. These measures often involve questions of economic values and trade-off. The question is whether these policies are making us better off or not. A society that is concerned with the well-being of its citizens should make changes in resources allocations only if what is gained by the change is worth more in terms of individuals' welfare than what is given up by diverting resources (Freeman et al. 2014).

Policy decision-making might involve, besides effects on private goods, the valuation of changes in public good provision. Theses goods and services might be, however, difficult to assess monetarily because they are not properly regulated by markets due to externalities as well as their characteristics of non-excludability and non-rivalry (Freeman et al. 2014).

The values of the environment are typically distinguished into two categories: use values and nun-use values (Adamowicz et al. 1998). Use values encompass direct consumptive values such as the value of timber, fish, recreation or other resources that ecosystems provide. Non-use values are economic values arising from a change in environmental quality that is not reflected in any observable behavior (Adamowicz et al. 1998). Such values may arise, for instance, from the desire to bequeath certain environmental resources to one's heirs or future generations, a sense of stewardship or responsibility for preserving certain features of natural resources or a desire to preserve options for future use (Freeman et al. 2014).

Several methods have been proposed to value environmental goods and services. They may be distinguished by the data source they are based on: Revealed preference (RP) data comes from observations of individuals acting in real-world settings. Valuation methods relying on RP data include the travel cost demand model or hedonic property value approach (Freeman et al. 2014). Methods based on stated preference (SP) data use surveys to elicit what an individual would be willing to pay or accept for a change in an environmental amenity (Adamowicz et al. 1998). Within the SP category two major approaches have been suggested: contingent valuation (CV), and discrete choice experiments (DCEs). While RP techniques rely on what people do, the SP approaches capture preferences based on what people say. Economists have therefore for a long time favored the former over the later (Kling et al. 2012). Nevertheless, CV and DCEs are in many instances the only available techniques that can be employed. Such situations arise when:

1. Data, such as housing transactions necessary for the hedonic pricing method, is not available or hard to obtain (Freeman et al. 2014)

2. Use-values associated with changes that fall outside the range of current markets or observed conditions (Johnston et al. 2017)
3. And, particularly, the valuation of non-use values. Here, only SP approaches can be employed (Freeman et al. 2014, Kling et al. 2012, Adamowicz et al. 1998).

Given this background, the development of valid and reliable SP methods becomes crucial.

The first application of the CV approach is attributed to Davis (1963) who used a questionnaire to value recreational hunting opportunities in the Maine woods (Poe 2016). After this work, numerous of other studies have been conducted using CV. In the last years the field of SP methods has been moving away from CV to more general DCEs (Desvousges et al. 2016).

DCEs, which may be seen as a generalization of the CV method since they use two or more alternatives which are described by attributes, arose from conjoint analysis which is often used in marketing (Adamowicz et al. 1998). First applications of DCEs, most of them in transportation research, date back to the 1980s (Louviere and Woodworth 1983, Louviere and Hensher 1982). Among the first who implemented a DCE in an environmental context were Adamowicz et al. (1994), who studied recreation site choices in Alberta, Canada.

Although applications of SP surveys in general and DCEs in particular have been increasing, the use of these methods is still highly controversial (e.g. Haab et al. 2013, Hausman et al. 2012, Kling et al. 2012). The critique of SP methods is often associated with the hypothetical nature and the possibility of deriving biased welfare estimates, referred to as hypothetical bias, which has been dominated the discussion on SP techniques over the last at least 20 years (Desvousges et al. 2016, Adamowicz 1994). Biases may further arise from not accounting for deviations from the general assumptions of “rational” consumer theory underlying discrete choice analyses, i.e. stable, fully compensatory, discounted expected utility maximization (Johnston et al. 2017). Individuals’ decisions that are not fully consistent with these assumptions have been termed behaviorally “anomalous”. Behavioral anomalies have been documented in many experimental and real-market settings with individuals employing a variety of decision heuristics or mental shortcuts (Kahneman et al. 1991, Kahneman 2003). Both, behavioral anomalies in the context of SP questionnaires - also referred to as (behavioral) response anomalies -, and heuristic-based decision-making have also been documented in the context of SP approaches (e.g. Johnson et al. 2017, Poe 2016). Moreover, it has been argued that factors such as unfamiliarity and complex information might make environmental valuation even more prone to anomalies (Carlsson 2010).

The design of SP surveys and DCEs involves several issues including attribute selection, description of alternatives, the design of the DCE as well as warm-up and auxiliary survey questions, etc. The central question, then, is whether response anomalies are triggered by the design of the SP survey, and how possible biases should be accounted for in the modeling process. Hensher (2006, p. 874) states in this regard "…*what matters is not whether different designs require different attribute (and information) processing strategies, but whether the stated choice design per se contributes to different behavioral responses and associated attribute valuations*." In their contemporary guidance for stated preference studies Johnston et al. (2017, p. 61), therefore, explicitly recommend that "*When prior research or pretesting indicates that undesirable response anomalies may be influential, data analysis should investigate these anomalies to determine whether they significantly affect SP responses*." Since DCEs are researcher-designed and based on the idea of constructing a hypothetical market, they give the researcher the flexibility to explicitly test for response anomalies by systematically manipulating the survey instrument (Carlson 2010).

This dissertation aims at analyzing response anomalies in DCEs for environmental valuation. An essential step in designing a DCE is to determine the design dimensions (number of choice sets, number of alternatives, etc.) and thereby the complexity of the DCE. In the light of widespread evidence on humans' limited ability in processing information (e.g. Simon 1956), influences of choice task complexity are analyzed by systematically varying the design dimensions of the DCE. It is investigated whether respondents choose less consistently or pre-maturely abandon the survey when complexity increases. Taking this as a point of departure, task complexity is then linked to systematic deviations from rational consumer theory regarding two decision heuristics: attribute non-attendance (ANA) and the status quo effect.

Besides ignoring attributes, respondents might also exclude alternatives from their choice set by using attribute thresholds (Swait 2001). Based on this, the use of attribute thresholds (cut-offs) is analyzed. Anomalies may further result from respondents not revealing their true preferences (Johnston et al. 2017). It is argued that only if a respondent perceives that her or his choices have a non-zero chance of influencing policy, truthful preference revelation can be expected (Carson et al. 2014, Carson and Groves 2007). Given this background, it is investigated whether preferences are impacted by information about the consequences of the survey and perceived survey consequentiality.

The remainder of this introduction is structured as follows. I begin by laying out the theoretical framework of DCEs and briefly present validity concepts related to SP approaches for environmental valuation. Then, I discuss the empirical evidence which calls into question the basic assumptions of SP elicitation. Here, I first present some key findings from behavioral economics and then discuss its relevant manifestations in DCEs and discrete choice models. Lastly, I develop the research questions by narrowing the research focus and identifying research gaps, present the empirical strategy as well as the outline of this dissertation.

1.1 Foundation and Validity of DCEs and Discrete Choice Modeling

In DCEs respondents are asked to choose their preferred alternative (option) from two or more hypothetical alternatives. Usually, participants in a DCE are presented with a sequence of such choice sets (also termed choice tasks or choice situations). Each alternative consists of attributes that can adopt different attribute levels. In order to be able to calculate welfare measures such as willingness to pay (WTP) or willingness to accept (WTA), one attribute is the cost of providing the good or service (Louviere et al. 2000). In DCEs for environmental valuation one alternative of the choice set usually describes the current situation or the status quo (Johnston et al. 2017). The observed choices are then used to derive preferences and welfare measures by means of discrete choice models. In the following, the foundations of DCEs and discrete choice modeling are presented. Then, the validity of DCEs is briefly discussed.

1.1.1 Foundations of Choice Modeling

The concept of economic valuation has its foundation in rational consumer theory. Here, the basic assumption is that the purpose of economic activity is to increase the well-being of individuals. Each individual’s welfare depends on his or her consumption of private as well as non-market goods and service (Freeman et al. 2014).

It is assumed that individuals’ consumption of a good or service depends on his or her preferences, which can be represented by means of a utility function. Rational consumer theory makes certain assumptions about these preferences and utility functions. Assumptions are typically made with respect to the ordering of preferences, namely transitivity, convexity and non-satiation (Freeman et al. 2014). Furthermore, preferences are assumed to be stable and fully known by individuals. Thus, they are invariant to the context in which they are revealed or stated (Freeman et al. 2014, Leong and Hensher 2012).

Another core assumption with is of particular importance in terms of deriving welfare measures such as WTP is substitutability (Freeman et al. 2014). Here, it is assumed that a decrease in one component of good (alternative) can by compensated by an increase in an element of another alternative making the individual indifferent between options. Preferences fulfilling this property are termed compensatory (Johnston et al. 2017). It is further commonly assumed that the decision maker chooses the alternative that maximizes her or his utility subject to a budget constraint thereby solving an optimization problem (Freeman et al. 2014).

Another important assumption, which is typically made in SP analysis, is that all respondents truthfully answer the SP question being asked (McNair et al. 2012).

Rational consumer theory has been extended by Lancaster (1966) and his “New Approach to Consumer Theory”. Based on Lancaster (1966) goods and services possess, or give rise to, multiple characteristics in fixed proportions and that it is these characteristics, not the goods themselves, on which the consumer’s preferences are exercised.

Another important step toward modern empirical discrete choice analysis was made with respect to individuals’ utility functions. Random utility models (RUMs) have been introduced in the 20th century by Marshak (1960). RUMs assume that an individual can choose from a finite number of alternatives with each alternative providing her or him with a certain level of utility. In line with utility maximization it is assumed that the decision maker chooses the alternative that yields him or her the highest level of utility. This utility is known to the decision maker, but not to the researcher, who only observes some of the attributes (Train 2009). Since the researcher does not know all components of individuals’ utility function, utility functions in RUMs are most commonly composed of a linear, additive separable deterministic part as well as a stochastic error term. This error term accounts for all the aspects the researcher is unable to observe (Train 2009).

In 1974, McFadden showed that if the error terms are independently and identically distributed following a Type I Extreme Value distribution, choice probabilities can be obtained following the logit formula - the Multinomial Logit model. However, this model imposes certain restrictions on individuals’ choices such as the independence of irrelevant alternatives (Freeman et al. 2014). By assuming other distributions for the error term, a large family of less restrictive choice models can be obtained.

Based Lancaster (1966) and by introducing prices and income by means of an “indirect” utility function, different measures of welfare can be derived by means of discrete choice models. Compensating variation measures or WTP / WTA may be calculated by using the marginal rate of substitution between a non-price and a price attribute (Hess et al. 2018, Freeman et al. 2014).

1.1.2 Validity of Stated Preference Surveys

Since their introduction to non-market valuation SP approaches have been subject to a controversial debate on their validity. Particularly two events sparked this controversy: The

Exxon Valdez oil spill in 1989 and the in Alaska's Prince William Sound and the 2010 Deepwater Horizon oil rig in the Gulf of Mexico. These oil spills led to large damages to the environment. Based on the economic assessment of these damages researcher came to very different conclusions regarding the usefulness of SP approaches. They ranged from "hopeless" and "useless" (Hausman 2012) to "…*some carefully constructed number based on stated preference analysis is now likely to be more useful than no number*…" (Kling et al. 2012, p. 23).

The critique to SP analysis is, in general, reflected in the discussion of reliability and validity. While reliability, which is not the focus of this dissertation, refers to the minimization of variability in estimates, validity refers to the minimization of bias in estimates (Johnston et al. 2017). Validity is often discussed within the following concepts:

1. Criterion validity: Do stated preferences estimates match real payments (Kling et al. 2012)? Tests of this concept usually address the possibility of obtaining welfare estimates SP surveys, which are biased upwards (hypothetical bias). Although empirical evidence is mixed, meta-analyses by Harrison and Ruttsröum (2008), Murphey et al. (2006) and List and Gallet (2001) tend to conclude that estimates obtained from CV and DCEs indeed exceed those from real payments. However, it has been argued this bias might be the result of the design of the SP study such as inconsequential treatments and a lack of incentive compatibility (Carson et al. 2014).
2. Convergent validity: Are stated and revealed preference estimates the same (Kling et al. 2012)? The answer to this question is mostly "yes" (Carson et al. 2014). Even when differences between stated and revealed preferences are detected, they are usually the result of common economic phenomena (Kling et al. 2012).
3. Content validity: Does the measure adequately cover the construct's domain, does the estimate arise from the best study design practices (Kling et al. 2012)? Tests of this validity concept may involve many different issues. Content validity can be achieved if design and implementation of the SP study follow best-practice guidelines such as those proposed by Johnston et al. (2017).
4. Construct validity: Are stated preference estimates consistent with theoretical predictions? Tests of validity may also consider many different issues. Examples are: Attribute non-attendance, protest responses… (Johnston et al. 2017).

1.2 Questioning the Assumptions of Stated Preference Elicitation

Section 1.1 has shown the SP elicitation is based on several assumptions such as compensatory, utility maximization behavior and truthful preference revelation. This section, first, provides a selective summary of empirical evidence from behavioral economics which questions the assumptions of rational consumer theory. Based on this, empirical findings with respect to response anomalies in DCEs and discrete choice modeling are discussed.

1.2.1 Core Findings from Behavioral Economics

Over the last decades, several studies in economics and contributions from psychology have been calling into questions the basic assumptions of rational consumer theory. As early as 1956, Simon asked the questions how humans respond to a choice task when the conditions of neoclassical theory are not met. His research became from then on known as "bounded rationality". Bounded rationality recognizes that individuals have cognitive constraints with respect to limitations in knowledge, time and computational capacity. Based on this idea and in contrast to utility-maximization behavior, Simon (1955) developed a model that assumes satisficing rather than maximizing behavior. Based on the idea of bounded rationality, Heiner (1983), theoretically, related individuals' information processing limitations to decision behavior in the context of complex choices. He suggested that individuals make error or adopt simplifying strategies in complex choices (Swait and Adamowicz 2001).

After Simon's contribution numerous empirical examples have been presented in which individuals in real-world examples as well as laboratory experiments have shown deviations from rational consumer theory with humans applying a variety of decision heuristics or mental shortcuts (Kahneman 2003). Decision heuristics recognize that when humans make decisions, some information is ignored in order to "*making decisions more quickly, frugally, and/or accurately than more complex methods*" (Gigerenzer and Gaismaier 2011, p. 454).

The empirical research on behavioral anomalies has also led to the development of new theoretical frameworks such as Prospect Theory (Tversky and Kahneman 1992, Kahneman and Tversky 1979).

The non-exhaustive list of examples of systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational consumer behavior (Kahneman 2003) and decision heuristics include:

- Status quo bias / status quo effect: Individuals have a strong tendency to remain at the status quo because the disadvantages of leaving it loom larger than advantages (Samuelson and Zeckhauser 1988, Kahneman et al. 1991).
- Endowment effect: People often demand much more to give up an object than they would be willing to pay to acquire it (Thaler 1980, Kahneman et al. 1991).

- Preference reversal: Individuals may reverse their preferences (Lichtenstein and Slovic 1971).
- Anchoring: Individuals decisions are affected by external cues / initial pieces of information (Tversky and Kahneman 1974).
- Satisficing: People do not maximize utility, they choose the alternative that meets their aspiration level (Simon 1955).
- Lexicography: Individuals make decisions by only evaluating aspects of each alternative (Tversky 1969).
- Elimination-by-aspects: Individuals view each alternative as a set of aspects. At each stage in the decision process, an aspect is selected and all the alternatives that do not include the selected aspect are eliminated. The process continues until all alternatives but one are eliminated (Tversky 1972).
- Preference construction: Preferences are constructed rather than merely revealed (Slovic 1995).

1.2.2 Anomalous Response Behavior in Stated Preference Surveys

The numerous examples of deviations from rational consumer theory have been echoed by research on anomalous response behavior in DCEs and discrete choice models. From a modeling point of view anomalies may manifest themselves in two different ways. First, RUM-based choice models may capture anomalous response behavior by means of the error term of the utility function. In the light of bounded rationality behavior that introduces “noise” would be in line with the passive bounded rationality model, which assumes that individuals continue to attend to all information in the choice set, but they increasingly make “mistakes” in processing that information. This would result in less consistent choices and, in turn, increases in the error variance (DeShazo and Fermo 2004, de Palma et al. 1994).

Second, deviations from the underlying assumptions may impact the utility function systematically. This behavior has been termed rational-adaptive (DeShazo and Fermo 2004).

Empirical research on anomalous response behavior in DCEs may at least be distinguished into three non-mutually exclusive categories (Johnston et al. 2017, Czajkowski et al. 2014):

- Information processing limitations
- Systematic patterns of respondents’ decision making
- Theoretical issues concerning SP surveys for environmental valuation / strategic misrepresentation

Within the first body of literature, the influence of choice task complexity has received considerable attention. In the DCE context choice task complexity is largely defined by the choice task properties, in particular, the design dimensionality of the DCE (Leong and Hensher 2012). The design dimensionality refers to the number of choice sets, the number of alternatives, the number of attributes, the number of attribute levels, and the level range. It is typically implied that choice designs with more items are more complex than those with less items (Hensher 2006, Swait and Adamowicz 2001). Other measures that have been used to capture choice task complexity relate to the structure and configuration of information in a DCE. Examples of such structural complexity measures include the number attribute level changes across alternatives, the number of trade-offs to be made by the person answering the choice tasks as well as a composite index (entropy) of choice task complexity, which captures complexity effects resulting from the similarity between alternatives (Zhang and Adamowicz 2011, DeShazo and Fermo 2002).

From the perspective of the researcher or practitioner, the presentation of multiple choice tasks per respondent, for instance, is preferred, and in some cases necessary, because it greatly increases the statistical efficiency (e.g. McNair et al. 2012). However, the complexity of the DCE has been observed to impact parameter and WTP estimates (Carlson et al. 1012, Zhang and Adamowicz 2011, Caussade et al. 2005).

Information processing limitations in DCEs are usually analyzed be means of the error variance where a higher error variance (also interpreted as choice consistency) is explained by effects of fatigue (or boredom) while a lower error variance is, amongst other things, attributed to learning or preference matching (Carlson et al. 2012, Day et al. 2012, Caussade et al. 2005). Learning effects may be further distinguished into institutional learning, which relates to the fact the most respondents in DCE have never participated in a SP survey, and value or preferences learning, which refers to the possibility that respondents discover of form their preferences while progressing through the sequence of choice tasks (Czajkowski et al. 2014).

Evidence on the relationship between each dimension and choice consistency is mixed, particularly with respect to the number of choice tasks. Here, results range from a systematic decreasing impact on the error variance pointing to learning (Czajkowski et al. 2014, Hess et al. 2012a), to increases in the error variance suggesting fatigue (e.g. Bradley and Daly 1994), to a quadratic relationship with the error variance first decreasing and then increasing with the choice set number (e.g. Caussade et al. (2005).

The number of alternatives has been mostly observed to impact error variance following a quadratic relationship, i.e. preference matching versus fatigue (e.g. DeShazo and Fermo 2002), while for the number of attributes research so far suggests a positive impact on the error variance (Caussade et al. 2005, DeShazo and Fermo 2002). Increases in the error variance have also been observed regarding entropy (Swait and Adamowicz 2001) as well as the correlation structure of attributes (DeShazo and Fermo 2002). Relatively little is still known on the influence of the number of attribute levels and the level range (see Caussade et al. 2005 for an exception).

Besides impacts on the error variance, complex choices may also result in the application of decision heuristics linking this body of literature to research on systematic patterns of respondents' decision making. When the DCE is complex respondents might adopt non-compensatory decision heuristics, i.e. they do not attend to all information of the choice set in order to simplify choices (Leong and Hensher 2012). One such heuristic could be to stay at the status quo, the status quo effect. In DCEs, status quo effects are usually identified through the number of choices of the status quo alternative reflected in the alternative-specific constant of the status quo. They are defined as "*the systematic inclination of respondents to display a different attitude towards status quo alternatives from those reserved to alternatives involving some change, over and beyond what can be captured by the variation of attributes' levels across alternatives*" (Scarpa et al. 2005, p. 250). Studies which related the status quo effect to choice task complexity include Boxall et al. (2009) who observed a positive relationship between status quo choices and the number of choice sets while Adamowicz et al. (2011) found a negative effect on status quo choices with respect to the number of alternatives presented on a choice set. Zhan and Adamowicz (2011) further observed status quo choices to be positively related to the number of trade-offs respondents have to make. The complexity of the DCE has also been observed to impact the attendance to attributes. Hensher (2006) found that the number of attributes ignored tends to increase as the number of levels of each attribute increases, as the range of each attribute narrows and as the number of alternatives decreases.

In general, ANA, i.e. whether respondents attend to all or ignore certain attributes, has received considerable research interest. Accumulating empirical evidence indicates that not all respondents attend to all attributes and that ANA can significantly bias WTP estimates (Campbell et al. 2011, Scarpa et al. 2009, Hensher et al. 2005). Non-compensatory decision heuristics such as ANA have been empirically identified directly from the discrete choice model (analytical approach) or by using additional information from the respondent obtained through supplementary survey questions (stated approach). Examples of the former approach are Hess et al. (2012b) and Daniel et al. (2018) who found that a considerable number of respondents choose according to elimination-by-aspects or lexicographic strategies. A prominent example for using stated information is Swait (2001) who proposed a model to account for attribute cut-offs, i.e. minimum (maximum) acceptable levels an individual sets for an attribute. The Swait (2001) model allows accounting for satisficing and elimination-by-aspects behavior and cut-off violation (Swait 2001). Empirical contributions that employed this model largely found that attribute cut-offs and cut-off violations play a role in respondents' decision making and thus may influence preferences (e.g. Ding et al. 2012, Bush et al. 2009).

Also within the body of literature analyzing systematic patterns of decision behavior, significant research effort has been dedicated to the stability of preferences in making choices with evidence suggesting that preferences may be constructed rather than revealed (Czajkowski et al. 2014). Particularly when faced with a sequence of choice tasks, preferences have been observed to be affected by the order of choice sets / ordering effects (Carlson et al. 2012, Day et al. 2012). Several explanations have been put forward for ordering anomalies including preference learning or anchoring, i.e. previous choice sets influence preferences in a sequence of choices (Hess et al. 2018, Day et al. 2012). If anchors are based on the first choice set, this effect has also been termed starting point bias (Day and Pinto Prades 2010). Anchoring and starting point bias have been widely documented in the discrete choice literature. Studies have shown that WTP when calculated from responses to follow-up questions is significantly lower than that calculated from first questions (Day and Pinto Prades 2010).

Another important phenomenon studied in DCEs is the gain-loss asymmetry and subsequent differences in welfare measures (WTP versus WTA). These asymmetries are related to the endowment effect, loss aversion and reference dependence. Evidence of the presents of these phenomena is widespread in the discrete choice literature (Hess et al. 2018) and the well-documented WTP/WTA-gap (see Horowitz and McConnell 2002 for a review) has also been reported in DCE studies (e.g. Bateman et al. 2009).

A third strand of literature investigates whether response anomalies can be attributed to non-truthful preference revelation. As mentioned in Section 1.1, SP approaches assume that respondents truthfully answer the questions being asked (McNair et al. 2012). Carson and Groves (2007) set out two conditions under which truthful preference revelation can be expected. First, the agent needs to care about what the outcomes of those actions might be. Second, the agent answering the survey needs to view his or her responses as potentially influencing decision-makers. Surveys satisfying these criteria are viewed as 'consequential', which is a crucial condition to achieve incentive compatibility (Carson and Groves 2007). In order to obtain information on consequentiality perceptions, participants are asked to state whether they think that their choices will be taking into consideration by policy makers (Vossler and Watson 2013). If a respondent perceives that there is a non-zero probability of influencing policy, truthful preference revelation can be expected. If, however, the likelihood of having an influence is perceived as zero, economic theory could not make any predictions. In this case individuals might, for instance, behave strategically by always choosing the most expensive alternative if they want the environmental good to be provided (Carson et al. 2014, Carson and Groves 2007).

Evidence so far suggests that SP responses equal real payments if there is a positive probability that responses to SP questions have real consequences. Otherwise, WTP estimates have been observed to be significantly different (Carson et al. 2014, Vossler and Watson 2013). Based on this the question arose how to best design the questionnaire to achieve non-zero consequentiality perception. Here, cheap talk scripts or consequentiality devices, respectively, which inform respondents about the consequences of the survey, have been proposed (Howard et al. 2017). Herriges et al. 2010) found that such a consequentiality device to be positively associated with the perceived degree of consequentiality.

1.3 Thesis Objectives and Outline

The last section has shown that response anomalies may influence choice outcomes and bias parameter and WTP estimates. This calls into question the validity – particularly criterion and construct validity - of DCEs, which aims at minimizing of bias in estimates (Johnston et al. 2017). The last section has also discussed a variety of different sources of response anomalies. Given this background, the following section develops the research questions to be addressed in this dissertation by narrowing the research focus and identifying research gaps. The empirical strategy to answer these research questions is presented subsequently. Finally, the outline of this dissertation is presented.

1.3.1 Development of Research Questions

This thesis is structured along five major research questions:

- Research question 1: How does choice task complexity influence survey drop-out rates, choice consistency and derived preferences?
- Research question 2: How does choice task complexity affect attribute non-attendance?
- Research question 3: How does choice task complexity influence the status quo effect and subsequently derived preferences?
- Research question 4: How do attribute cut-offs affect preferences?

- Research question 5: How does survey consequentiality influence preferences?

With respect to choice task complexity, rational consumer theory would predict that, independently of the complexity of the DCE, individuals are cognitively indefatigable examining all alternatives and all attributes across all choice sets in the same fully compensatory manner (Leong and Hensher 2012). Behind this background and in the light of the literature discussed in the last section, this thesis, first, seeks to investigate the influence of choice task complexity on different aspects of response behavior.

So far, evidence with respect to each dimension of choice task complexity tends to be mixed (e.g. the number of choice sets). In addition, there is only one series of studies (Rose et al. 2009, Hensher 2006, Caussade et al. 2005, Hensher 2004) in which the influence of the complexity of the DCE is investigated systematically. Systematically, here, means that the dimensions of complexity, i.e. the number of choice sets, the number of alternatives, the number of attributes, the number of attribute levels and the level range, are varied in a systematic manner (Hensher 2004). These contributions, however, were conducted in a transportation context. Here, the characteristics of the good under valuation differ to those in the environmental domain (e.g. respondents' experience and familiarity with the good). This thesis is intended to fill this research gap by systematically investigating impacts of choice task complexity in an environmental context.

A somewhat extreme impact of choice task complexity would be that respondents refuse to make choices at some point in the sequence of choice sets by dropping out of the survey. For participants completing the DCE, choices may become more random / choice consistency decreases when the information load increases due to fatigue effects. Based on this, research question 1 analyzes how choice task complexity influences survey drop-out rates, choice consistency and derived preferences. When the DCE becomes more complex, respondents might also not process all the available information of the choice set by applying non-compensatory decision heuristics (Leong and Hensher 2012). Making trade-offs is cognitively demanding and participants may cope with this demand by avoiding to make trade-offs (Payne et al. 1999).

A decision heuristic to cope with choice task complexity may be the exclusion of attributes from the decision making process. When choices are complex respondents may simplify the task by focusing on a subset of attributes that they consider to be most relevant or important. As already pointed out, ANA in DCEs has been intensively studied. However, research on the relation between task complexity and ANA is still scarce (see Hensher 2006 for an exception from transportation research). Research question 2, therefore, investigates how choice task complexity affects ANA.

Although rarely related to the dimensions of complexity (see Boxall et al. 2009, DeShazo and Fermo 2004 for two exceptions), another decision heuristic is to disproportionally sticking with the status quo alternative, i.e. the status quo effect (Payne et al. 1999). When complexity increases, uncertainty might also increase preventing individuals from making welfare-enhancing choices, for instance (DeShazo and Fermo 2004). As a consequence, participants retain what is known (a type of endowment effect) or to avoid the decision entirely by deciding not to choose (Boxall et al. 2009).

Behind this background research question 3 studies how choice task complexity affects the status quo effect and subsequently derived preferences.

Apart from the complexity of the DCE, respondents' unwillingness to make trade-offs may also be associated with decision heuristics imposing minimum (maximum) requirements for attributes (Swait 2001). Alternatives that do not satisfy these requirements are excluded from the choice sets. The preferred alternative is then chosen from the remaining choice set (Ding et al. 2012). Passed research suggests that individuals indeed employ attribute cut-offs in DCEs. However, most of this evidence stems from transportation and food applications while, to the best of my knowledge, only Bush et al. (2009) analyzed cut-offs in an environmental context. Furthermore, there is, as far as I know, no study that systematically analyzes whether prior cut-off elicitation itself affects preferences. As already discussed, there is widely documented evidence that preferences elicited in SP surveys may be constructed rather than revealed (Czajkowski et al. 2014). Thus, preferences might be affected by cut-off questions asked prior to the DCE. Research question 4 sheds light on how attribute cut-offs affect preferences.

Lastly and as discussed in Section 1.2, response anomalies may also be triggered by individuals who perceive their survey responses to be inconsequential. Although research on survey consequentiality has been increasing, there is still relatively little known about the relationship between preferences and consequentiality perceptions. In addition, the determinants that influence perceived consequentiality such as information about the consequences of the survey has so far received little or no attention. Research question 5 addresses these issues by studying the influence of survey consequentiality on preferences.

1.3.2 Empirical Approach

The research questions developed above will be addressed by means of two specially tailored SP surveys. A common approach to systematically test for influences of different survey designs is the use of treatments or split samples. Both surveys follow this approach by implementing four and, respectively, 16 treatments:

Research question 1, 2 and 3 are addressed by using data from a nation-wide online survey on land-use changes in Germany. In order to investigate influences of choice task complexity, the study followed an approach with 16 split samples in which the number of choice sets, the number of alternatives, the number of attributes, the number of attribute levels and the level range were varied systematically. The variation of these design dimensions largely followed a Design of Designs approach introduced by Hensher (2004). Since several studies published in recent years have shown that respondents can cope with a fairly large number of choice sets, the number of choice tasks was slightly increased compared to Hensher (2004). Also, the number of attributes was higher compared to Hensher (2004).

Participants were randomly assigned to one of the 16 treatments. In order for choice sequence effects not to be confounded with choice task effect, the position of each choice sets within each design was also randomized. On each choice set respondents were asked to choose their preferred alternative from a status quo (current situation) and, depending on the design, two, three or four generic alternatives.

The attributes of the DCE were “share of forest”, “land conversion” and different biodiversity attributes, which were based on an indicator of bird populations. This indicator allows one to vary the number of attributes by splitting up bird populations into different parts of the landscape. In addition “contribution to a landscape fund” was used as a payment vehicle.

Besides the DCE the questionnaire included a number of auxiliary survey questions. To identify respondents’ attendance to attributes (research question 2) stated and analytical approaches have been suggested (see also Section 1.2.2). In order to be able to employ both methods, respondents indicated whether they attended to the attributes shown on the choice sets after they had completed the experiment. At the beginning of the survey, socio-demographic characteristics were asked. In order to control for possible influences of the perceived status quo in answering research question 3, participants indicated their perception of the current situation regarding the attributes implemented in the corresponding treatment.

The survey was conducted in 2012. At each stage of the questionnaire respondents could only abandon the survey by closing their web browser. In this case the drop-out position was recorded. In total, 1,684 respondents completed the questionnaire while 263 participants dropped out of the survey within or after the DCE.

Research question 4 and 5 are addressed by analyzing data from a nation-wide online survey on the development of renewable energies in Germany. The survey was conducted in 2013 and 3390 usable interviews were collected. The questionnaire included a DCE with six choice sets (the position of each choice sets was randomized) in which respondents had the choice between four labeled alternatives (the order of the first three alternatives was also randomized):

- Electricity from wind energy (wind farms)
- Electricity from solar energy (solar fields)
- Electricity from biomass (biogas power stations)
- I do not care about the type of renewable energy generation (you will not have any influence on the type of renewable energy which will be developed in the 10 km surroundings of your place of residence)

Each alternative was described by six attributes: The minimum distance of renewable energy facilities to the edge of town, the size, and number of renewable facilities, protection of the landscape view, whether new high-voltage transmission lines are built overhead or underground, and the change of the electricity bill. Again, the questionnaire comprised additional questions for answering the research questions. To address research question 4 participants were pleased to state their minimum (maximum) requirements (cut-offs) for the minimum distance, the number of renewable facilities, protection of the landscape view and whether new high-voltage transmission lines are built overhead or underground, prior to the DCE. Cut-offs were only elicited from half of the sample to which respondents were randomly assigned.

Also prior to the DCE, participants were informed about the consequences of the survey as well as the institution to be responsible for future renewable energies expansion. This information was varied systematically using four treatments to which respondents were also randomly assigned. Following the DCE, respondents stated their perceived consequences of the survey. This information was then used to address research question 5. Again, socio-demographic characteristics such as age, gender and education were collected at the beginning of the survey.

More details on study design, survey implementation as well as the modeling approach may be found in the corresponding research article.

1.3.3 Thesis Outline

This thesis consists of seven chapters which all contribute to answering the above raised research questions. Chapter 2, 3, 4 and 6 represent articles published in peer-reviewed journals.

Chapter 5 is a manuscript / conference paper presented at the 5th International Choice Modeling conference 2017 as well as the 6th World Congress of Environmental and Resource Economists. The author of this dissertation it the second author of Chapter 2 and Chapter 3 and the first author of Chapter 4 and Chapter 6 as well as the single author of Chapter 5.

The remainder of this thesis is structured as follows:

- Chapter 2: The Influence of Design Dimensions on Stated Choices in an Environmental Context
- Chapter 3: Stated and inferred attribute non-attendance in a design of designs approach
- Chapter 4: Uncovering context-induced status quo effects in choice experiments
- Chapter 5: Do attribute cut-offs make a difference? The effects of eliciting and incorporating cut-off values in choice models
- Chapter 6: Stated preferences towards renewable energy alternatives in Germany – do the consequentiality of the survey and trust in institutions matter?

This dissertation closes with Chapter 7 – the synthesis – where the main results are summarized and discussed, implications for researchers and practitioners are presented, and future research is highlighted.

1.4 References

Adamowicz, W., Dupont, D., Krupnick, A., Zhang, J. 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. Journal of Environmental Economics and Management 61(2), 213–226.

Adamowicz, W. 1998. Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. American Journal of Agricultural Economics 80(1), 64–75.

Adamowicz, W.L., Louviere, J., Williams, M. 1994. Combining Stated and Revealed Preference Methods for Valuing Environmental Amenities. Journal of Environmental Economics and Management 26, 271–292.

Bateman, I.J., Day, B.H., Jones, A.P., Jude, S. 2009. Reducing gain–loss asymmetry: A virtual reality choice experiment valuing land use change. Journal of Environmental Economics and Management 59(3), 271–285.

Boxall, P., Moon, A., Adamowicz, W.L. 2009. Complexity in Decision Making: The Effect of Task Structure and Context on Participant Response. Australian Journal of Agricultural and Resource Economics 53(4), 503–519.

Bradley, M., Daly, A. 1994. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21(2), 167–184.

Bush, G., Colombo, S., Hanley, N. 2009. Should all Choices Count? Using the Cut-Offs Approach to Edit Responses in a Choice Experiment. Journal of Environmental and Resource Economics 44(3), 397–414.

Campbell, D., Hensher, D.A., Scarpa, R. 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. Journal of Environmental Planning and Management 54(8), 1061–1076.

Carlsson, F., Mørkbak, M.R., Olsen, S.B. 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling 5(2), 19–37.

Carlson, F. 2010. Design of Stated Preference Surveys: Is There More to Learn from Behavioral Economics? Journal of Environmental and Resource Economics 46(2), 167–177.

Carson, R. T., Groves, T., List, J.A., 2014. Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. Journal of the Association of Environmental and Resource Economists 1, 171–207.

Carson, R. T., Groves, T. 2007. Incentive and informational properties of preference questions. Journal of Environmental and Resource Economics 37(1), 181–210.

Caussade, S., Ortúzar, J., Rizzi, L.I., Hensher, D.A. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Ressearch Part B: Methodological 39(7), 621–640.

Czajkowssk, N., Giergiczny, M., Greene, W.H. 2014. Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. Land Economics 90, 324–351.

Daniel, A.M., Persson, L., Sandorf, E.D. 2018. Accounting for elimination-by-aspects strategies and demand management in electricity contract choice. Energy Economics 73, 80–90.

Davis, R.K. 1963. The Value of Outdoor Recreation: An Economic Study of the Maine Woods. Ph.D. Dissertation, Harvard University.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63(1), 73–91.

DeShazo, J.R., Fermo, G. 2004. Implications of rationally-adaptive pre-choice behavior for the design and estimation of choice models. Working paper, School of Public Policy and Social Research, UCLA.

DeShazo, J.R., Fermo, G. 2002. Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management 44, 123–143.

de Palma, A., Myers, G.M., Papageorgiou, Y.Y. 1994. Rational choice under an imperfect ability to choose. American Economic Review 84, 419–440.

Desvousges, W., Mathews, K., Train, K. 2016. From Curious to Pragmatically Curious: Comment on "From Hopeless to Curious? Thoughts on Hausman's 'Dubious to Hopeless' Critique of Contingent Valuation†". Applied Economic Perspectives and Policy 38(1), 174–182.

Ding, Y., Veeman, M.M., Adamowicz, W.L. 2012. The influence of attribute cut-offs and consumers' choices of a functional food. European Review of Agricultural Economics 39(5), 745–769.

Freeman, A.M., Herriges, J.A., Kling, K.L. 2014. The Measurement of Environmental and resource Values: Theory and Methods. London: Taylor & Francis.

Gigerenzer, G., Gaissmaier, W. 2011. Heuristic Decision Making. Annual Review of Psychology 62, 451–483.

Haab, T.C., Interis, M.G., Petrolia, D.R. 2013. From Hopeless to Curious? Thoughts on Hausman's “Dubious to Hopeless” Critique of Contingent Valuation. Journal of Applied Economic Perspectives and Policy, 35(4), 593–612.

Harrison, G.W., Rutström, E.E. 2008. Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In: Handbook of Experimental Economics Results 1(5), edited by Plott, C.R., Smith, V.L. North-Holland: New York.

Hausman, J., 2012. Contingent valuation: from dubious to hopeless. Journal of Economic Perspectives 26(4) 43–56.

Heiner, R. A. 1983. The origin of predictable behavior. American Economic Review 73, 560–595.

Hensher, D.A. 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics, 21, 861-878.

Hensher, D.A., Rose, J.M., Greene, W.H. 2005. The implications on willingness to pay of respondents ignoring specific attributes. Transportation 32, 203–222.

Hensher, D.A. 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. Journal of Transport Economics and Policy 38(3), 425–446.

Herriges, J., Kling, C., Liu, C.-C., Tobias, J. 2010. What are The Consequences of Consequentiality?” Journal of Environmental Economics and Management 59: 67–81.

Hess, S., Daly, A., Batley, R. 2018. Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory and Decisio 4(2), 181–204.

Hess, S., Hensher, D.A., Daly, A. 2012a. Not bored yet – Revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: Policy and Practice 46(3), 626-44.

Hess, S., Stathopoulos, A., Daly, A. 2012b. Allowing for heterogeneous decision rules in discrete choice models: An approach and four case studies. Transportation 39(3), 565-591.

Horowitz, J.K., McConnell, K.E. 2002. A Review of WTA/WTP Studies. Journal of Environmental Economics and Management 44(3), 426–447.

Howard, B.E., Roe, E., Nisbet, C., Martin, J.C. 2017. Hypothetical Bias Mitigation Techniques in Choice Experiments: Do Cheap Talk and Honesty Priming Effects Fade with Repeated

Choices? Journal of the Association of Environmental and Resource Economists 4(2) 543–573.

Johnston, R.J., Boyle, K.J., Adamowicz, W.L, Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C.A. 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists 4(2), 319–405.

Kahneman, D., Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47(2), 263–292.

Kahneman, D., Knetsch, J.L., Thaler, R.H. 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. Journal of Economic Perspectives 5(1), 193–206.

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

Kling, C.L., Phaneuf, D.J., Zhao, J. 2012. From Exxon to BP: Has some number become better than no number. Journal of Economic Perspectives 26(4), 3–26.

Lancaster, K.J. 1966. A New Approach to Consumer Theory. Journal of Political Economy 74(1), 132–57.

Leong, W., Hensher, D.A. 2012 Embedding decision heuristics in discrete choice models: a review. Transport Reviews: A Transnational Transdisciplinary Journal 32(3), 313–331.

Lichtstein, S., Slovic, P. 1971. Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology 89(1), 46–55.

List, J.A., Gallet, C.A. 2001. What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values? Journal of Environmental and Resource Economics 20(3), 241–54.

Louviere, J.J., Hensher D.A., Swait, J. 2000. Stated choice methods: analysis and application. Cambridge University Press: Cambridge.

Louviere, J.J., Woodworth, G. 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of Marketing Research 20, 350–357.

Louviere, J.J., Hensher, D.A. 1982. On the Design and Analysis of Simulated Choice or Allocation Experiments in Travel Choice Modelling. Transportation Research Record 890(1), 11–17.

Marschak, J. 1960. Binary Choice Constraint on Random Utility Indications. In: Stanford Symposium on Mathematical Methods in the Social Sciences, 312–319, edited by Arrow, K. Stanford: Stanford University Press.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics, edited by Zarembka P., 139–145. Academic Press: New York.

McNair, B.J., Hensher, D.A., Bennett, J. 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: A probabilistic decision process model. Journal of Environment and Resource Economics 51(4), 599–616.

Murphy, J.J., Allen, P.G., Stevens T.H., Weatherheard, D. 2005. A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. Journal of Environmental and Resource Economics 30(3), 313–425.

Payne, J.W., Bettman, J.R., Schkade, D.A. 1999. Measuring Constructed Preferences: Towards a Building Code. Journal of Risk and Uncertainty 19(1), 243–270.

Poe, G.L. 2016. Behavioral Anomalies in Contingent Values and Actual Choices. Agricultural and Resource Economics Review 44(2), 246–269.

Rose, J., Hensher, D.A. Caussade, S., de Dios Ortúzar, J., Jou, R.-C. 2009. Identifying Differences in Willingness to Pay due to Dimensionality in Stated Choice Experiments: A Cross Country

Analysis. Journal of Transport Geography 17(1) 21–29.

Samuelsson, W., Zeckhauser, R. 1988. Status Quo Bias in Decision Making. Journal of Risk and Uncertainty, 1, 7–59.

Scarpa, R., Gilbride, T.J., Campbell, D., Hensher, D.A. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. European Review of Agricultural Economics, 36(2), 151–174.

Scarpa, R., Ferrini, S., Willis, K.G. 2005. Performance of Error Component Models for Status-Quo Effects in Choice Experiments. In: Applications of Simulation Methods in Environmental and Resource Economics, ed. Scarpa, R. Alberini, A. Netherlands: Springer.

Simon, H.A. 1955. Rational choice and the structure of the environment Psychological Review 63(1956), 129–138.

Slovic, P. 1995. The Construction of Preference. American Psychologist 50, 364–371.

Swait, J. 2001. A non-compensatory choice model incorporating attribute cut-offs. Transportation Research Part B: Methodological 35(10), 903–928.

Swait, J., Adamowicz, W. 2001. Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. Organizational Behavior and Human Decision Processes 86(2), 141–167.

Thaler, R. 1980. Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization 1(1), 39–60.

Train, K. 2009. Discrete choice methods with simulation. Cambridge University Press: Cambridge.

Tversky, A., Kahneman, D. 1992. Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and Uncertainty 5(4), 297–323.

Tversky, A., Kahenman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. Science 185(4157), 1124–1131.

Tversky, A. 1972. Elimination by aspects: A theory of choice. Psychological Review 79(4), 281–299.

Tversky, A. 1969. Intransitivity of preferences. Psychological Review 76(1), 31–48.

Vossler, C.A., Watson, S.B. 2013. Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. Journal of Economic Behavior and Organization 86, 137-147.

Zhang, J., Adamowicz, W.L. 2011. Unraveling the Choice Format Effect: A Context-Dependent Random Utility Model. Land Economics 87(4), 730–743.

2 The Influence of Design Dimensions on Stated Choices in an Environmental Context

Jürgen Meyerhoff[a], Malte Oehlmann[a], Priska Weller[b]

Abstract

Discrete choice experiments are increasingly used in the context of environmental valuation. However, there is still little known about the influence of the complexity of the choice task on model outcomes. In this paper we investigate task complexity in terms of the design dimensionality of the choice experiment by systematically varying the number of choice sets, alternatives, attributes, and levels as well as the level range. We largely follow a Design of Designs approach originally introduced in transportation. First, we analyse the influence of the design dimensionality on participants' dropout behaviour finding that the probability to drop-out of the survey is influenced by socio-demographic characteristics and increases with the number of choice sets, attributes as well as with designs having five alternatives. Second, we investigate the impact of the design dimensions on stated choices by estimating a multinomial logit model, and heteroskedastic logit models. Results show that the error term variance is influenced by socio-demographic characteristics as well as by all design dimensions. Moreover, we find that accounting for the impact of the design dimension on the error variance does not significantly change willingness to pay estimates.

Keywords: Choice complexity · Design of designs · Stated choice experiment · Error variance · Land use changes

Published in: Environmental and Resource Economics (2015), 61: 385–407, https://doi.org/10.1007/s10640-014-9797-5

[a] Institute for Landscape and Environmental Planning, Technische Universität Berlin, Straße des 17. Juni 145, 10623 Berlin, Germany

[b] Johann Heinrich von Thünen Institute, Institute for International Forestry and Forest Economics, Leuschnerstraße 91, 21031 Hamburg, Germany

2.1 Introduction

Discrete choice experiments (CE) are widely used to elicit consumer preferences in various applications including marketing, health economics, transportation and environmental valuation. CEs usually consist of a series of hypothetical scenarios or choice sets, each including two or more alternatives. These alternatives are described by attributes whose levels are varied systematically. For each choice set respondents are then asked to rank the alternatives, choose best and worse or select their preferred alternative (Louviere et al. 2000).

The number of choice sets, alternatives, and attributes as well as the number of levels and their range are the design dimensions of a CE. These are to be specified by the researcher. The dimensionality of a CE can vary significantly across different studies. For instance, in the majority of CE studies respondents are offered four to ten choice sets and two to four alternatives. However, there are occasions when participants are asked to assess up to 26 choice tasks (Czajkowski et al. 2014) and choose among 12 or more alternatives (Chung et al. 2011).

Assuming neoclassic economic theory, which suggests that decision makers can perfectly and consistently process information, the design dimensionality should not influence choice outcomes (Swait and Adamowicz 2001). As a consequence, the complexity of the CE, the ability of the individual to make complex decisions and the effect of the choice context on decision strategies are often not considered in statistical models. However, many authors have highlighted the need to account for the limited ability of individuals to process complex information. Heiner (1983), for instance, argued that choice complexity can influence choice consistency and the more complex the choice task, the higher the gap between individual's cognitive ability and cognitive demand. This result leads to a trade-off to be made by the researcher. On the one hand, one might be tempted to increase task complexity, for instance, to increase the number of observations or to gather more information by increasing the number of attributes. On the other hand, this comes at the cost of higher cognitive burden and the risk of significantly distorted model estimates (DeShazo and Fermo 2002). As a result, there is so far no agreement in the literature on the optimal task complexity in CEs.

Task complexity can be viewed as a part of the unobserved factors influencing choice outcomes. Generally, unobserved influences are classified as omitted variables, measurement errors in the observed attributes and alternatives, true task complexity that imposes variation in cognitive difficulty, uncertainty attributable to many sources such as stimulus ambiguity and beliefs about future states and peer impacts (Hensher 2006).

Our study investigates the issue of true task complexity by varying five design dimensions. The research is largely motivated by a series of studies investigating the influence of five design dimensions - the number of choice sets, alternatives, attributes, attribute levels and their range - on CE outcomes in the context of transportation (Hensher 2004; Caussade et al. 2005; Hensher 2006; Rose et al. 2009). In these studies, 16 different treatments were used which were generated by a Design of Designs (DoD) approach originally introduced by Hensher (2004). The attributes used considered different travel times and costs. The present case study is based on a nation-wide online survey carried out in Germany about land use changes. The survey was completed in December 2012 and incorporated aspects of land conversion, such as the share of forest and different biodiversity attributes. Similar to Hensher (2004) and Caussade et al. (2005) we used 16 different split samples.

To our knowledge, this is the first study applying a DoD approach in environmental valuation by systematically varying five design dimensions. We analyse the influence of the five design dimensions in terms of two aspects. First, we investigate the relationship between task complexity and participants' drop-out behaviour by using descriptive statistics and by specifying a binary logit model. Second, we estimate a joint multinomial logit model (MNL) and heteroskedastic logit models (HL) with the scale parameter specified as a function of the design dimensions and socio-demographic characteristics of the respondents. This allows us to determine the impact of the five design dimensions on the error term variance which is inversely related to the scale parameter. Willingness-to-pay (WTP) estimates are subsequently obtained from both models and compared to each other.

In the remainder Section 2 outlines previous literature and the hypotheses to be tested. Next, Section 3 presents the modelling approach before detailing the study design and implementation (Section 4) and presenting results (Section 5). Finally, Section 6 discusses the results.

2.2 Literature Review and Hypotheses

The influence of task complexity in stated CEs has been investigated in several studies. Typically, complexity issues have been analysed in the context of health economics (Ryan and Wordsworth 2000; Ratcliffe and Longworth 2002; Bech et al. 2011), marketing research (Dellaert et al. 1999; Dellaert et al. 2012) and transportation (Hensher et al. 2001; Hensher 2004; Caussade et al. 2005; Hensher 2006; Rose et al. 2009). The research has focused largely on dimensionality influences on the error variance or scale parameter considering the effects of fatigue and learning (Dellaert et al. 1999; DeShazo and Fermo 2002; Arentze et al. 2003; Caussade et al. 2005; Rolfe and Bennett 2009; Chung et al. 2011; Czajkowski et al. 2012) and on WTP estimates (Ryan and Wordsworth 2000; Hensher 2004; Hensher 2006; Bech et al. 2011; McNair et al. 2011). Effects on attribute weights (Arentze et al. 2003), response and completion rate (Hensher et al. 2001; Bech et al. 2011; Louviere et al. 2013), decision time (Dellaert et al. 2012) and perceived choice certainty (Rose et al. 2009; Brouwer et al. 2010; Bech et al. 2011) have also been analysed in existing research. Regarding the impacts on the error variance and in the light of studies published in recent years, we largely follow the hypotheses that were put forward by Caussade et al. (2005) and compare their results to our findings in the context of environmental valuation.

2.2.1 Task Complexity and Drop-out Rates

To our knowledge only very few studies have investigated the impact of task complexity on participants' drop-out behaviour in discrete CEs. Two studies explicitly investigate the relationship between the number of choice sets and the response rate. Hensher et al. (2001) found little differences in response rates in mail surveys with varying numbers of choice tasks, while Bech et al. (2011) found no significant influence in an online survey. However, it has to be noted that response and drop-out rates must be distinguished from each other. Whereas the former gives the ratio between those respondents who completed the questionnaire and those contacted, the latter shows the percentage of those participants who prematurely abandoned the questionnaire (Vicente and Reis 2010). So far the most comprehensive study concerning completion rates was presented by Louviere et al. (2013). They used a web panel to elicit preferences for two consumer goods, pizza and flights. Across interviews the authors systematically varied the number of choice sets (16 or 32), the number of alternatives (3 to 5), the number of attributes (6 to 12) and the type of statistical design ("Street and Burgess" or "Kuhfeld" design). They found that, among other things, completion rates are relatively unaffected by presenting more choice sets to respondents, but decline when the sets comprise more alternatives. In contrast, studies carried out in other disciplines concluded that the survey length has a significant influence on drop-out rates (see for example Vicente and Reis 2010). Furthermore, a study by Galesic (2006) shows that the lower the experienced burden, the lower the risk to drop out. We thus expect that both, the number of choice sets and the number of alternatives, are positively related to drop-out rates although the results presented by Louviere et al. (2013) indicate only a weak effect.

Finally, for the number of attributes as well as for the number of levels and their range, it may be argued that the number of comparisons to be made increases with both, more attributes and a higher number of levels. Moreover, comparisons between alternatives might be easier to assess for attribute levels which have a narrow range (Caussade et al. 2005). As a result, we also hypothesise a positive relationship between these design dimensions and drop-out rates.

2.2.2 Task Complexity and Error Variance

Number of Choice Sets: So far the most frequently investigated design dimension is probably the number of choice sets. However, there is still no consensus in the literature about impacts on the error variance. On the one hand, Hensher et al. (2001) found little evidence for fatigue effects for even 32 choice sets; Czajkowski et al. (2012) support this inference. On the other hand, Bradley and Daly (1994) found the error variance to increase as the number of choice sets grew. Caussade et al. (2005) and Chung et al. (2011) observed a U-shaped relationship with the error variance decreasing up to a threshold (nine/ten sets in Caussade et al. 2005; six sets in Chung et al. 2011) and then increasing beyond this. A similar pattern was found by Bech et al. (2011). They argue that the error variance initially decreases due to learning effects while subsequent fatigue effects cause the error variance to increase. Also, findings by Scarpa et al. (2011) point in this direction, who based their analysis on rank-ordered data elicited with the best-worst approach. In their study each respondent faced 16 choice sets. Scale increased gradually from the first to the eleventh set and declined for the last three sets. Following these arguments we expect a U-shaped pattern for the relation between the number of choice sets and scale.

Number of Alternatives: The evidence and subsequent conclusions about the impact of the number of alternatives tend to be mixed. Arentze et al. (2003) found no effects on the error variance when distinguishing between designs with two and three alternatives. However, Caussade et al. (2005) as well as Chung et al. (2011) found a U-shaped pattern with the lowest error variance for five and four alternatives respectively. A similar pattern emerged in the study conducted by DeShazo and Fermo (2002), who argued that the initial decrease of the error variance results from a better match of preferences while the increase at the later stage is caused by a more complex choice. Our hypothesis is thus that a U-shaped relationship between the number of alternatives and the error variance exists.

Number of Attributes: There is clear evidence suggesting that an increase in the number of attributes results in an increase in the error variance. Caussade el al. (2005) found that the number of attributes has a strong detrimental effect on a participant's ability to choose which contributes to a higher error variance. Similar inferences were drawn by DeShazo and Fermo (2002) and Arentze et al. (2003). As the information load to be processed by the respondent grows with the number of attributes, we also expect a positive relationship between the error variance and the number of alternatives.

Number of Attribute Levels and Level Range: Based on the findings from Dellaert et al. (1999) and Caussade et al. (2005) and following the same arguments presented above regarding the influences on drop-out rates, we expect a positive relationship between the number of attribute levels and the error variance. Moreover, we expect an increase in the error variance with a wider level range.

2.3 Econometric Approach

Random utility theory assumes the modeller does not possess complete information concerning the individual decision maker (subscript n). Thus, individual preferences are the sum of a systematic (V) and a random (ε) component

$$U_{ni} = V_{ni}(x_{ni}\beta) + \varepsilon_{ni} \tag{1}$$

where U_{ni} is the true but unobservable utility associated with alternative i out of a set of available alternatives, V_{ni} is the measurable or deterministic part which itself is a function of the attributes (x_{ni}), β is a vector of coefficients reflecting the desirability of the attributes, and ε_{ni} is a random term with a zero mean. This error term represents attributes and characteristics unknown to the researcher, measurement error and/or taste heterogeneity among respondents. Selection of one alternative over another implies that the utility (U_{ni}) of that alternative is greater than the utility of the other alternative:

$$P(i) = \Pr ob(V_i + \varepsilon_i > V_j + \varepsilon_j) \quad \forall\, j \in C, j \neq i \tag{2}$$

Assuming that the error components are distributed independently and identically (IID) following a type 1 extreme value distribution, one gets the multinomial logit (MNL) model where the probability of individual *n* choosing alternative *i* takes the form:

$$P_{ni} = \frac{\exp(\mu V_{ni})}{\sum_{j \in C} \exp(\mu' V_{nj})} \tag{3}$$

where μ is a scale parameter which is commonly normalised to 1 in practical applications for any one data set as it cannot be identified separately from the vector of parameters. The scale parameter is inversely proportional to the error variance σ_ε^2:

$$\mu = \frac{\pi}{\sqrt{6\sigma_\varepsilon^2}} \tag{4}$$

The assumption of a constant error variance across individuals has been questioned and a heteroskedastic logit model was suggested as an alternative (HL; e.g., Swait and Louviere, 1993). Here the scale parameter is no longer a constant term as it allows for unequal variances across unobserved components from two or more data sources. Whether the variation in scale across data sources can be explained by factors such as respondent characteristics or the design dimensions of the choice sets is investigated using the following HL expression (Caussade et al. 2005; DeShazo and Fermo 2002):

$$P_{ni} = \frac{\exp(\mu_n V_{ni})}{\sum_{j \in C} \exp(\mu_n V_{nj})}$$

(5)

where μ_n = $\exp(\gamma' Z_n)$ with Z_n a vector of respondent specific characteristics including the design dimensions a respondent is randomly assigned to and γ' a vector of parameters indicating the influence of those characteristics on the error variance (Hole 2006). The exponential form ensures a positive scale factor.

In both heteroskedastic logit models estimated in this study the scale parameter is specified as a function of the five design dimensions based on the DoD approach (number of choice sets, alternatives, attributes, attribute levels and the level ranges) as well as the respondent characteristics age, gender, and education. The two models differ with respect to the specification of the number of choice sets. In the first model (HL1) we consider this effect as linear and quadratic, while in the second model (HL2) we use dummy coded choice set numbers (CS_t) in order to capture scale dynamics along the sequence of choice sets where CS_1=0. The parameters γ' and β' are jointly estimated via maximum likelihood using the Stata program *clogithet* (Hole 2006).

2.4 Study Design and Implementation

2.4.1 Study Design

Our study design largely follows the design master plan introduced by Hensher (2004). Several studies published in recent years have shown that respondents can cope with a fairly large number of choice situations. Thus, we slightly adapted the design master plan of Hensher (2004) and use 6, 12, 18 and 24 instead of 6, 9, 12 and 15 choice sets. In contrast to the Hensher (2004) study we also increased the number of attributes from four to seven instead of three to six. All other dimensions were kept equal to those of Hensher (2004).

In order to obtain 16 treatments we generated 16 different generic designs using Ngene software. Unlike Hensher (2004) and Caussade et al. (2005), who employed a D-efficient experimental design, we made use of the C-error as a design criterion in each treatment as it minimises the variance of the WTP estimates (Scarpa and Rose 2008). Our adaptation of the design master plan can be seen in Table 1.

Table 1: Adaptation of the design master plan

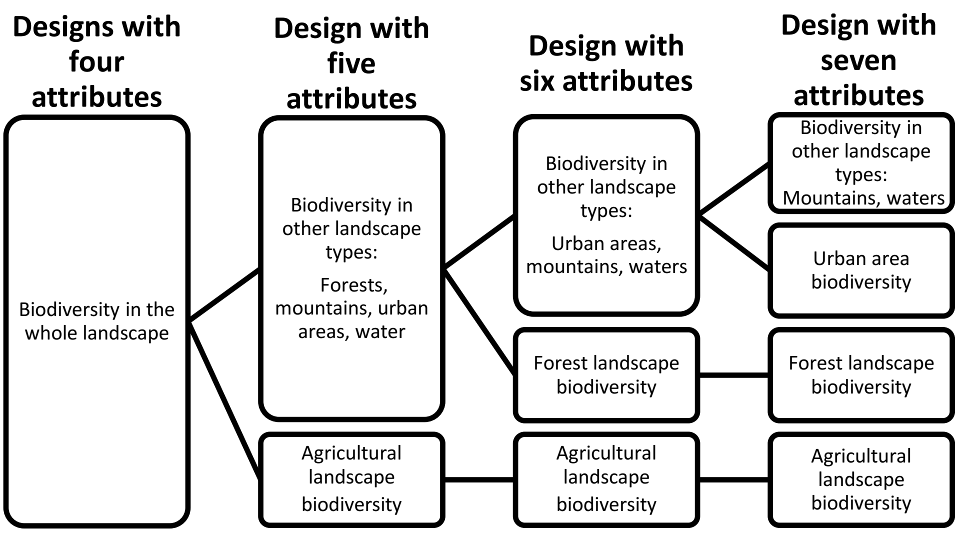
Design	Sets	Alternatives	Attributes	Levels	Range
1	24	4	5	3	Base
2	18	4	5	4	+20%
3	24	3	6	2	+20%
4	12	3	6	4	Base
5	6	3	4	3	+20%
6	24	3	4	4	-20%
7	6	4	7	2	-20%
8	12	5	4	4	+20%
9	24	5	4	4	Base
10	6	5	7	3	+20%
11	6	4	6	4	-20%
12	12	5	5	2	-20%
13	18	4	7	2	Base
14	18	3	4	3	-20%
15	12	3	5	2	Base
16	18	5	6	3	-20%

The attributes in the present study deal with land use changes. For all non-price attributes the number of levels and the level range varied due to the design master plan. We distinguish between three different groups of attributes. Firstly, the attributes "share of forest" and "land conversion" were included in all treatments.

The second attribute, for example, aims at the rate on converting mainly farmland into sealed land as a result of building infrastructure. The different level values were expressed in percentage changes compared to the current state. Secondly, different biodiversity attributes were used which were based on an indicator of bird populations. This indicator was developed as part of an indicator system for sustainable development in Germany (BMU 2010). The indicator can be split into bird populations in different parts of the landscape, e.g. "birds in the whole landscape" equals "birds in agricultural landscapes" plus "birds in other landscapes". We varied the number of attributes across treatments following the methodology used by Hensher (2004), who used different types of travel time and cost. This allowed us to aggregate and disaggregate the biodiversity attribute as a combination of already existing attributes. In doing so we were able to systematically account for the influence of the number of attributes on model outcomes (Hensher 2004).

Figure 1 illustrates the split of biodiversity attributes across different designs. The levels were expressed as indicator values. Respondents were informed that an indicator of 100 or more means the landscape type is a good habitat for a variety of species.

Figure 1: Split of the biodiversity attribute



Thirdly, "contribution to a landscape fund" was utilised as the payment vehicle. This attribute was presented in all designs and the number of levels and level ranges were constant over all treatments. We are aware that the payment vehicle used in our study might cause problems concerning the issue of incentive compatibility. However, we decided not to use taxes due to previous experiences with protest responses in Germany. Table 2 summarises the attributes used in our study and their associated levels for a base design with two levels and a base level range (Designs 13 and 15). The attribute levels for the other designs are reported in Appendix 1.

Table 2: Attributes and their levels for the base design

Attribute	Description	Levels	Unit
Share of forest	Percentage changes in the share of forest	-25, +25	%
Land conversion	Percentage changes in the land conversion	-50, +50	%
Bio_whole	Biodiversity in the whole landscape including all landscape types	70, 100	indicator score
Bio_agrar	Agricultural landscape biodiversity	65, 100	indicator score
Bio_forest	Forest landscape biodiversity	80, 100	indicator score
Bio_urban	Urban area biodiversity	60, 100	indicator score
Bio_other1	Biodiversity in other landscape types: Forests, urban areas, mountains, waters	75, 100	indicator score
Bio_other2	Biodiversity in other landscape types: Urban areas, mountains, waters	60, 100	indicator score
Bio_other3	Biodiversity in other landscape types: Mountains, waters	75, 100	indicator score
Cost	Contribution to a landscape fund per year	10, 25, 50, 80, 110, 150	€

As it can be seen from Table 1, the number of alternatives varies from three to five including the status quo alternative, which was defined as the current situation with the cost attribute of zero. The attribute levels for the status quo alternative were "as today" as we asked respondents to assess land use changes for an area within 15 km of their residence without knowing the shape of the landscape in each respondent's surrounding. This is another factor that distinguishes our study from Hensher (2004). Further differences are summarised in Table 3.

Table 3: Differences across complexity studies

Characteristic	Hensher (2004)*	Present Study
Application	Transportation	Environmental valuation
Number of choice sets	6, 9, 12, 15	6, 12, 18, 24
Number of attributes	3 – 6	4 – 7
Survey mode	Computer aided personal interview (CAPI)	Online survey
Experimental design	D-efficient	C-efficient
Status quo	Current route	Current situation
Region	Sydney	Germany (nationwide)
Aim of assessment	Route changes	Land use changes around 15 km of residence
Variation of attribute numbers	Based on different types of travel time and of costs	Based on Biodiversity in different landscapes types
Payment vehicle	Travel cost	Contribution to a landscape fund
Number of observations	8,020	90,354

* see also Caussade et al. (2005), Rose et al. (2009), Chintakayala et al. (2010)

2.4.2 Survey Implementation

A nation-wide online survey was conducted between December 7th and December 21st 2012 using a panel from a survey company. When participants entered the survey, they were randomly assigned to one of the 16 designs. The questionnaire started by asking respondents socio-demographic questions concerning date of birth, gender and education. We did this at the initial stage of the questionnaire in order to be able to include socio-demographic characteristics in the drop-out analysis. Then, participants were asked several ‘warm-up’ questions in order to introduce them to the topic and to obtain their assessment of the perceived current state of landscape characteristics within 15 km of their residence. Before presenting the first choice set, participants were given an instruction page with information on the CE as well as the attribute descriptions. These descriptions varied across treatments with four, five, six and seven attributes. In order for the choice sequence effect not to be confounded with the choice task effect, we randomised the sequence of choice tasks within each design. So, the position of each choice set differed across respondents. For each choice set respondents were asked to assess land use changes for an area within 15 km of their residence by choosing their preferred alternative. Figure 2 shows an example choice set. For the biodiversity attributes the value of today's indicator score was presented on each choice card. Additionally, respondents were informed that these values could vary across regions in Germany. For the attributes "share of forest" and "land conversion" we do not know the present situation in the 15km surrounding of each respondents. Thus we did not provide any values for the current situation and used relative changes as attribute levels.

Figure 2: Example choice set with four alternatives and seven attributes

If only the following options were available for the future development of the landscape within a radius of up to 15 kilometres around your place of residence, which one would you choose?
If you live in a large city, please consider the surrounding area of the city.

	Option 1	Option 2	Option 3	Option 4
Share of forest in the landscape	20% more	20% less	20% more	as today
Land conversion	40% more	40% more	40% more	as today
Biodiversity in agricultural landscapes (Today 65 scores)	76 scores	89 scores	89 scores	as today
Biodiversity in forest landscapes (Today 80 scores)	95 scores	95 scores	95 scores	as today
Biodiversity in other landscape types (Today 65 scores)	89 scores	89 scores	89 scores	as today
Biodiversity in urban areas (Today 60 scores)	90 scores	90 scores	70 scores	as today
Financial contribution to the landscape fund per year	80 €	50 €	80 €	0 €
I choose	○	○	○	○

After finishing the CE several follow-up questions were asked including perceived choice certainty and subjective attendance to attributes. As respondents progressed through the questionnaire, they saw a bar that indicated what fraction of the questionnaire had been answered. According to the number of choice sets respondents saw slower progression of the bar when they were assigned to a design with more choice sets. However, respondents were not informed in advance about how many choice sets they would face during the interview.

At each stage of the questionnaire respondents could only abandon the survey by closing their web browser. In this case the drop-out position was recorded. In total, 2122 interviews were collected with 1684 (79.36%) participants completing the whole questionnaire and 438 participants only partially completing it. For the group of respondents who quitted before the CE we do not have sufficient data as many respondents dropped out on the initial screens before any socio-demographic characteristics were requested. Therefore, we do not include this group in subsequent analysis. The average interview length was measured to be 23 minutes and the response rate was 29.49% (corresponding to the completed interviews relative to persons contacted). Table 4 reports basic socio-demographic characteristics for the sample including those respondents who started to answer the CE (completed plus drop-outs) as well as the sample used to estimate the choice models (completed interviews). Education refers to the highest educational level achieved. Across the 16 designs we could not find any statistically significant differences regarding the socio-demographic characteristics for both samples reported in Table 4.

Table 4: Socio-demographic characteristics

Variable		Sample including drop outs* mean (SD)	Sample completed interviews mean (SD)
Age in years		42.75 (13.69)	42.34 (13.59)
Gender (male = 1)		0.50 (0.50)	0.48 (0.50)
Highest educational level	Compulsory secondary school (9 years schooling)	0.07 (0.24)	0.07 (0.24)
	Middle-school degree (10 schooling years)	0.28 (0.45)	0.27 (0.44)
	High-school degree (13 schooling years)	0.26 (0.44)	0.27 (0.44)
	University degree	0.39 (0.49)	0.39 (0.49)
Number of respondents		1,947	1,684

Note: Middle-school is used here as an equivalent of the German “Realschule”. *This sample does not comprise those 175 respondents who left the survey before the CE started, SD = standard deviation

2.5 Results

2.5.1 The Influence of the Design Dimensionality on Drop-out Rates

As a first empirical illustration, Table 5 depicts the drop-out rates according to the stage of completion of the questionnaire. Around 8.2% of the participants dropped out before starting to answer the CE. The highest drop-out rate is within the CE (10.8%), while only 1.6% quit the survey after finishing the CE. The relatively high drop-out rate in the first part of the questionnaire has also been observed in studies carried out in other areas of research. Hoerger

(2010) also found 10.0% of the participants to drop out of the survey instantaneously. This behaviour may be explained by a low interest in the topic. Among others, Galesic (2006) found that the lower the respondent’s interest in the topic of the survey, the higher the drop-out rate. The authors asked respondents at the end of each of the 20 question blocks, which contained questions on a specific topic, how interesting they found the questions in the preceding block. For respondents who dropped out at any later point in the survey levels of interest were lower at previous blocks.

Table 5: Position of drop-outs

Drop-out position	Frequency	%
Before CE	175	8.25
Within CE	229	10.79
After CE	34	1.60
Completed	1684	79.36
Total	2122	100.00

Table 6 presents the 16 treatments with their corresponding design dimensions as well as the number of interviews and the number of drop-outs per design. We exclude those respondents who quit the survey before starting to answer the CE as all questions were identical across designs up to this stage. In total 263 respondents dropped out of the survey within and after the choice task section of the interview. The designs with the highest drop-out rates are the designs 9, 3 and 1. Each has 24 choice sets with all other design dimensions varying (three to five alternatives, four to six attributes, etc.). The three designs with the next highest drop-out rates all have 18 choice sets. At the other end of the spectrum we have the designs 2, 15, 5 with the lowest drop-out rates. All these designs have a different number of choice sets, but none has 24 choice sets. Thus, there is already some indication that drop-out rates are associated with the number of choice sets.

Table 6: Design-dependent drop-outs within and after the choice experiment

Design	Drop-outs			Design Dimensions				
	In %	Number	Complete interviews	Sets	Alternatives	Attributes	Levels	Range
9	25.93	21	81	24	5	4	4	Base
3	25.00	31	124	24	3	6	2	20%
1	21.18	18	85	24	4	5	3	Base
13	20.35	23	113	18	4	7	2	Base
14	20.24	17	84	18	3	4	3	-20%
16	20.24	17	84	18	5	6	3	-20%
10	18.54	28	151	6	5	7	3	20%
8	16.46	13	79	12	5	4	4	20%
6	15.85	13	82	24	3	4	4	-20%
4	15.00	12	80	12	3	6	4	Base
12	12.35	10	81	12	5	5	2	-20%
11	11.36	10	88	6	4	6	4	-20%
7	11.26	25	222	6	4	7	2	-20%
2	10.98	9	82	18	4	5	4	20%
15	6.94	10	144	12	3	5	2	Base
5	5.77	6	104	6	3	4	3	20%
Total	**15.62**	**263**	**1684**					

In order to analyse the relationship between the five design dimensions and drop-out rates in detail, we specify a binary logit model with the dependent variable being zero if a participant completed the survey and one if the respondent dropped out after starting to answer the CE. The results are presented in Table 7. As expected the number of choice sets has a highly significant positive impact on the probability to drop out. The same pattern is observed for the number of attributes. Regarding the number of alternatives, only the dummy variable for five alternatives significantly influences the probability to drop out. This effect is, however, only significant at the 10% level. For the attribute levels and the level range we find no significant effect. So, we could not reject our null hypothesis of no effect. A linear regression on the influence of the design dimensions on the number of respondents who dropped of the survey confirms these findings. The results also show that the number of choice sets as well as the number of attributes positively influence the number of respondents who drop out (see Appendix 2 for the linear regression results).

Since we also expected socio-demographic characteristics to be possible explanations for the probability to drop out, we included age, gender and education as further independent variables. As shown in Table 7 all but the second age group have a significantly higher probability to drop out compared to respondents being between 18 and 30 years old. Also the dummy for being male is positive and highly significant. With respect to education all parameters have a negative sign indicating that higher education leads to less drop-outs compared to the reference group. However, only the parameter for high-school degree is statistically significant. Finally, none of the interactions among design dimensions we incorporated in the model are statistically significant at the 5% level.

Table 7: Results of the binary logit model on drop-outs

Dimension for the probability to drop out	Coefficient	(|t-Value|)
Number of choice sets	0.0556	4.80
Dummy 4 alternatives	-0.0428	0.22
Dummy 5 alternatives	0.3355	1.85
Number of attributes	0.2060	2.32
Narrow level range	0.0727	0.42
Wide level range	0.0856	0.46
Number of levels	0.0720	0.68
Age group 31 to 40 years	0.4075	1.86
Age group 41 to 50 years	0.1094	0.51
Age group 51 to 60 years	0.5079	2.33
Age group older than 60 years	0.4906	2.09
Gender male	0.6888	4.90
Middle-school degree	-0.1680	0.65
High-school degree	-0.6169	2.19
University degree	-0.1994	1.44
Constant	-4.5138	5.46
Log-likelihood null	-770.88	
Log-likelihood model	-734.69	
Pseudo-R^2	0.05	

Note: N = 1,947 including 263 respondents who dropped out within or after the CE

2.5.2 Drop-outs within the Choice Experiment

To conclude our analysis on drop-out rates, Table 8 shows the drop-out position within the CE. We observe the highest number of participants (123) to drop out within the first six choice sets. This figure corresponds to 53.7% of the drop-outs within the CE and a drop-out rate of 6.3%. At the other extreme, only 20 respondents or 8.7% of the drop-outs abandoned the survey between choice set 19 and set 24. The total drop-out rate at this stage of the CE is therefore only 4.4%. Therefore, respondents who do not agree with the format of a choice experiment seem to drop out early while those who agree are likely to stay even if they face a longer sequence of sets.

Table 8: Drop-out rates within the choice experiment

Position drop out	Number of presented choice sets				Total	Respondents in respective designs	Drop-out rate
	6 Sets	12 Sets	18 Sets	24 Sets			
Set 1 to 6	57 *24.9*	20 *8.7*	21 *9.2*	25 *10.9*	123 *53.7*	1947	*6.3*
Set 7 to 12		16 7.0	19 8.3	16 7.0	51 22.3	1191	4.3
Set 13 to 18			16 7.0	19 8.3	35 15.3	849	4.1
Set 19 to 24				20 8.7	20 8.7	454	4.4
	57 *24.9*	36 *15.7*	56 *24.5*	80 *34.9*	229 *100.0*		

Note: percentages in italics

2.5.3 Heteroskedastic Logit Models

We estimated three different models to investigate the association between the design dimensions and scale. First, we report the results from a simple MNL model in which the data from all split samples are pooled and only the choice attributes are incorporated. Next, estimates from a heteroskedastic logit model (HL1) with the scale parameter as a function of the design dimensions, interactions among them, and socio-demographic s characteristics are reported in Table 9. Another heteroskedastic logit model (HL2) investigating order effects along the sequence of choice sets is presented in Appendix 2.

In all three models the attribute parameters are deemed significant at a 1% level of significance and have the expected sign. On average respondents prefer a larger share of forests while at the same time preferring reduced land conversion in their surroundings. The biodiversity attributes all have positive signs indicating that respondents want to increase the levels of biodiversity as measured by the underlying scale. The model also contains an alternative specific constant (ASCsq) for the current situation. It is positively and statistically significant suggesting that, on average, respondents have a propensity to choose the current situation instead of one of the hypothetical alternatives describing future land use changes.

Table 9: Estimation results for the MNL and HL1 model

Variable	MNL		HL1	
	Coefficient	|t-Value|	Coefficient	|t-Value|
ASCsq	0.5114	19.04	0.2230	2.99
Share of forest	0.0167	35.57	0.0079	3.02
Land conversion	-0.0085	38.31	-0.0041	3.03
Bio_whole	0.0096	8.61	0.0053	2.95
Bio_agrar	0.0054	9.36	0.0025	2.81
Bio_forest	0.0059	6.75	0.0029	2.70
Bio_urban	0.0061	5.81	0.0032	2.63
Bio_other1	0.0076	6.73	0.0033	2.76
Bio_other2	0.0027	3.85	0.0012	2.18
Bio_other3	0.0052	5.81	0.0022	2.52
Cost	-0.0062	26.51	-0.0028	3.06
Covariates				
Position of choice set			0.0354	4.16
Squared position of choice set			-0.0008	2.45
Dummy 4 alternatives			0.2262	3.35
Dummy 5 alternatives			0.0731	1.16
Interaction 4 alternatives * number CE			-0,0139	2,57
Interaction 5 alternatives * number CE			-0,0027	0,50
Number of attributes			0.1046	1.78
Number of levels			0.3147	3.15
Interaction attributes*levels			-0.0658	3.44
Narrow level range			0.2046	5.54
Wide level range			-0.1042	2.69
Age31to40			-0.0121	0.31
Age41to50			-0.0812	2.16
Age51to60			0.0587	1.50
Age60plus			-0.1672	3.54
Gender male			-0.1279	4.95
Middle-school degree			-0.0018	0.03
High-school degree			0.0954	1.61
University degree			0.1449	2.57
Log-likelihood null	-31114.18			
Log-likelihood model	-27438.90		-27291.44	
Observations	90354			

The estimates from the HL1 model regarding the effect of the design dimensions, interactions among them, and the socio-demographics characteristics on scale are reported in the lower part of Table 9. This model accounts for the number of choice sets by incorporating the position of each choice set as a linear and quadratic effect. Note that the choice sets were always presented in a randomized order for each design. Similar to Caussade et al. (2005) we find evidence for an inverse U-shaped relationship between the number of choice sets and scale. The linear effect shows that an increasing number of choice sets affects scale positively indicating a learning effect. As the respondent becomes familiar with the choice tasks and moves along the sequence of choices, a decrease in variance (and hence an increase in the scale of the Gumbel error) is observed. This has been attributed to learning. At the same time, though, an opposite effect is observed that has been attributed to increase in respondent's fatigue and results in an increased error variance (and hence a lowered Gumbel error scale). Towards the end of the choice sequence the latter effect overrides the former, thereby producing an overall increase in error variance.

With respect to the number of alternatives, we can reject the null hypotheses by observing a U-shaped pattern with dummy variables specified for four and five alternatives. Designs with four alternatives have a higher scale parameter (lower error variance) when compared to the designs with three alternatives. This suggests a similar pattern as observed by Caussade et al. (2005). However, the coefficient for the variable indicating five alternatives per choice set is not statistically significant in our model. Regarding both the number of attributes and the number of levels we find significant impacts on the scale parameter, but the effect of the number of attributes is only significant at the 10% level. Our data also rejects the null hypothesis of no effect on scale due to the last of our design dimensions: the range of attribute levels. We observe the error variance to decrease for designs with a narrow range and to increase for designs with a wide range compared to the base. Choice consistency is promoted when level ranges are narrow as comparisons among alternatives are easier to carry out. This result is again in line with Caussade et al. (2005).

Testing whether the effect of one design dimension depends on another we found two statistically significant interactions. Firstly, the interaction between the choice set position and the number of alternatives reveals that scale increases until the twentieth choice set. Afterwards it remains at the same level (Figure 3). For choice sets with four alternatives, however, we find that after a significant increase in scale at the beginning of the sequence and a modest increase between the choice sets 2 and 12, scale decreases resulting in an increased error variance in the second half of the sequence. The interaction for five alternatives and the position of the choice set is, however, not statistically significant. Thus, designs with three or five alternatives do not cause different error variances. Secondly, the highly significant interaction effect between number of attributes and number of levels indicates that the impact of both design dimensions on scale is conditional. An increasing number of attributes or attribute levels, everything else remaining equal, decreases the scale effect.

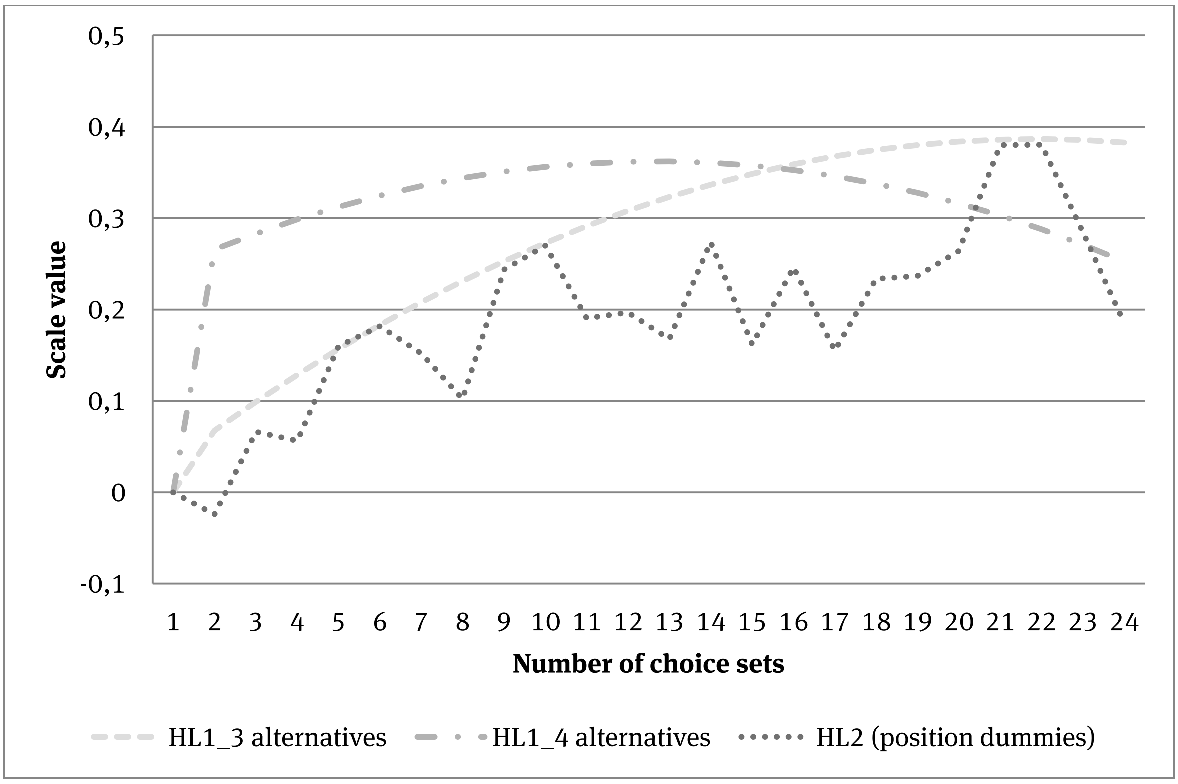
We find significant effects for all of the following socio-demographic covariates: age, gender, and education. Two of the age groups have a statistically significant lower scale parameter in contrast to the reference group of respondents age 18 to 30. Also, being male has a negative influence on the scale parameter while holding a university degree has a positive influence on scale compared to those with nine and less years schooling (the reference group). To compare both the MNL and HL1 model we use a log-likelihood ratio test. The test statistic is 147.46. Since the critical Chi-squared value is 33.41 (1% level with 17 degrees of freedom), we can reject the null hypothesis that the HL1 model is no better than the MNL model.

We now turn to the HL2 model accounting for order effects with dummies for each choice set except for the first (CS_2-CS_{24}; Appendix 2). Note that this model does not, unlike model HL1, simultaneously account for the number of alternatives per choice set due to the very high number of additional interactions required by such a model. The results suggest an order effect as the CS_2-CS_{24} coefficients are jointly significant. In contrast to other studies (Scarpa et al. 2011; Czajkowski et al. 2012; Carlsson et al. 2012), however, we do not find a clear learning effect on the first choice sets. The coefficients for the choice sets CS_2 to CS_4 are not statistically significant, the coefficient for CS_3 has a negative sign. The first dummy variable positively significant is related to CS_5. The highest values occur, on the other hand, for the sets CS_{21} and $CS_{22,}$ before the value decreases again. As the coefficient for the last choice set is not statistically significant, it might not be justified to argue that this drop in scale indicates fatigue.

The scale effects for both models are illustrated in Figure 3. Firstly, the dashed line (colored red) shows the scale effects due to model HL1 for the reference alternative with three alternatives per choice set. This line indicates a low but constantly rising scale value until CS_{21}-CS_{22} which afterwards remains at that level. Secondly, the line with dashes and dots (colored red) shows the scale effect for choice sets with four alternatives. For these designs we find a significantly positive effect at the beginning of the sequence. However, this effect decreases more and more when we move along the sequence of sets. After CS_{16} scale becomes lower than in the reference situation with three alternatives. This suggests that for designs with more than 16 sets it is advantageous to use three or five alternatives. Also, the latter does not show statistically significant effects on scale compared to designs with three alternatives. Finally, the dotted line (colored green) shows the scale effects for each choice set based on model HL2 (see Appendix 2). As these values capture the effect across designs with three, four and five alternatives, the values are always below the dashed blue line.

Overall, the results from both the HL1 and the HL2 models seem to support findings presented by Czajkowski et al. (2012; similar Hess et al. 2012). They found that scale rather remained on the same level after it has been increasing for the first eight to nine choice sets, thus, not clearly suggesting fatigue effects at the end of the sequence of choices. For designs with four alternatives, however, our results indicate fatigue suggesting that this effect differs across designs with a varying number of alternatives.

Figure 3: Scale effects along the order of choice sets



Note: In HL2 the dummies for the positions 2, 3, 4, 8, and 24 are not statistically significant at the 10% level

Table 10 reports the marginal WTP estimates for the MNL and the HL1 model. The confidence intervals were calculated using the Delta method. Note that the marginal WTP estimates refer to environmental changes in the area within 15 km of each respondent's residence. Starting with the estimates based on the MNL model, respondents are, on average, willing to pay 2.71 € per year for a one percent increase in the share of forests, but would experience a disutility of 1.38 € per year for a one percent growth in land conversion. The biodiversity attributes, which were aggregated and disaggregated across designs, delivered results as expected. Respondents are willing to pay more for the aggregated attribute "whole landscape biodiversity" (Bio_whole), which is 1.55 € per year for an increase in the indicator score, than for attributes at a lower aggregation level such a forest landscape biodiversity (0.95 € per year for an improvement of a score) or agricultural landscape biodiversity (0.88 € per year for an improvement of a score). Moreover, the WTP for Bio_other1, which includes biodiversity in forests, urban areas, mountains and waters, is higher than Bio_other2 considering biodiversity in mountains, urban areas and waters. In contrast, the WTP for Bio_other3 is close to the WTP value for Bio_other1 although it comprises less habitats.

Table 10: Marginal willingness to pay estimates for attributes in Euro per year

Attribute	MNL		HL1		Diff MNL – HL
	mWTP €	95%-CI	mWTP €	95%-CI	One-sided significance $mWTP_{MNL}$ - $mWTP_{HL1}$
Share of forest	2.71	2.43 / 2.99	2.85	2.55 / 3.15	0.38
Land conversion	-1.38	-1.50 / -1.27	-1.47	-1.60 / -1.34	0.33
Bio_whole	1.55	1.17 / 1.94	1.90	1.49 / 2.30	0.15
Bio_agrar	0.88	0.68 / 1.09	0.91	0.68 / 1.14	0.49
Bio_forest	0.95	0.66 / 1.25	1.04	0.69 / 1.39	0.49
Bio_urban	0.98	0.64 / 1.32	1.15	0.74 / 1.56	0.28
Bio_other1	1.24	0.86 / 1.62	1.19	0.79 / 1.60	0.42
Bio_other2	0.44	0.21 / 0.67	0.43	0.16 / 0.70	0.44
Bio_other3	0.86	0.56 / 1.15	0.79	0.44 / 1.15	0.37

Note: The one-sided p-values for the difference between both WTP estimates result from 2500 bootstrap replications; CI =confidence interval.

The absolute WTP estimates differ between the MNL and the HL1 models, but the point estimates are each time included in the interval of the other model for the same attribute. This indicates that the WTP estimates are not statistically different. Additionally, we directly bootstrapped the difference between both WTP estimates for each attribute using 2500 repetitions. In order to determine the approximate one-sided significance of the difference we follow the percentile approach and calculate the proportion of negative values of the difference between the marginal WTP estimates from both models (see Poe et al. 1997). The calculated values (last column Table 10), resulting from 2500 replications, support the conclusion that the difference between both models is not statistically significant. Thus, taking into account the impact of the design dimension and socio-demographic characteristics on the error variance does not result in significantly different WTP estimates.

2.6 Discussion

In this study we have analysed the impact of the number of choice sets, the number of alternatives in each choice set, the number of attributes as well as the associated number and range of levels on drop-out rates and the error variance. To our knowledge it is not only the first study employing the Design of Designs approach introduced by Hensher (2004) in environmental valuation, but also the first attempt to investigate the relationship between task complexity in discrete CE and participants' drop-outs in this field.

With respect to drop-out rates, we find that the probability to abandon the survey significantly increases with the number of choice sets and the number of attributes. Furthermore, designs with five alternatives show a higher drop-out rate compared to those with only three alternatives. All other design dimensions do not significantly influence the probability to quit the survey. Additionally, older respondents as well as men are more likely to drop out. Among those who dropped out while answering the choice tasks, it is noteworthy that the majority of the respondents abandoned the CE within the first six choice questions. One reason for this could be that people did not like the choice format and thus decided not to proceed. During focus groups we have repeatedly discovered that some respondents do not like to make comparisons among bundles of attributes, but would prefer to rate each attribute separately. Another reason might be that people realised that choosing among the alternatives on a choice set also affects payments to the landscape fund and that quitting the survey at this stage was motivated by protest votes. However, we do not conclusively know the reasons why respondents abandoned the survey; we refrained from sending those respondents debriefing questions as the survey company expected very low response rates for this kind of questions. The results overall suggest that if only drop-out rates in an online survey are of concern, fewer choice sets and attributes as well as less than five alternatives should be presented. However, Louviere et al. (2013) argue that due to the cost of statistical information, i.e., the survey costs per interview, the potentially negative effect of designs with a higher dimensionality on drop-out rates is overcompensated by the additional information gained. Applying their calculations (see Louviere et al. 2013: 28) to our data we can confirm this finding. The design with the lowest drop-out rate (Design 5: 6 choice sets, 3 alternatives, 4 attributes) would result in 1131 data records per 100 individuals starting the survey (and 94 finishing it). The design with the highest drop-out rate (Design 9: 24 choice sets, 5 alternatives, 4 attributes) would result in 7111 data records per 100 individuals starting to answer the survey (and 74 finishing it). Clearly, the design with a higher dimensionality would provide more information although the completion rate is significantly lower. Generally, drop-out rates and the reasons why respondents drop-out of a survey have not found much attention. This is particularly surprising for online surveys taking into account how easily this can be investigated using paradata. Whether drop-out rates depend on survey mode is another interesting question. In a computer assisted personal interview (CAPI), as used by

Czajkowski et al. (2014), for example, respondents might on average be more motivated or might feel more obliged to go through a longer sequence of choice sets.

With respect to influences of the design dimensions on the error variance, we mainly find the same results as Caussade et al. (2005). For all five dimensions we reject our null hypothesis of no relationship between the design dimensionality and the error variance. However, there are some differences compared to the findings by Caussade et al. (2005) in transportation. The first is that fatigue effects are not as clearly present as in their study. For designs with three and five alternatives, scale does not decrease as rapidly as in their study or as presented by other authors (e.g., Scarpa et al. 2011). In contrast, for designs with four alternatives we firstly observe a significant increase followed by a decreasing scale value. An explanation for this could be that there is an easier choice on sets with an odd number of alternatives. Overall, our findings are rather in line with results presented by Czajkowski et al. (2014) and Hess et al. (2012) who could not find strong evidence for fatigue effects due to the number of choice sets.

The second key difference is that overall no large increase in choice certainty is observable at the first choice sets in the sequence; only for designs with four alternatives we notice a stronger increase. This raises the question whether the first choices should be used as a valid expression of preferences or whether they should be seen solely as warm-up exercises. That choice certainty does not increase strongly at the beginning also raises the question whether instruction choice sets, as suggested for example by Carlsson et al. (2012), would counteract this successfully. The third difference worth emphasizing is that in our study also two interactions among design dimensions are statistically significant while Caussade et al. (2005) could not discover any significant interaction. As already mentioned, the position of the choice set in the sequence of choices and the number of alternatives interact. Designs with four alternatives perform better at the beginning of the sequence but within the course of that sequence the scale value significantly decreases indicating that this format is less suitable for longer sequences of choice sets. The second highly significant interaction effect is between the number of attributes and the number of attribute levels. A low number of attributes or levels has a positive effect on scale that decreases and turns into a negative effect when the number of attribute levels, all else remaining equal, increases.

For all attributes considered in our study the WTP estimates are calculated based on the MNL and the HL1 model. The WTP estimates are higher for an aggregated biodiversity attribute than for biodiversity attributes at a lower level of aggregation, except in one case where the WTP estimate for Bio_other3 is higher than for Bio_other2. This pattern points to valid WTP estimates. Comparing the estimates from the MNL and the HL1 model does not reveal significantly different values. Accounting for the impact of the dimensionality on scale differences does therefore not change the main findings regarding the WTP measures. This finding is again in line with the results presented by Caussade et al. (2005).

Generally, our results confirm those findings presented in the literature that indicate that a higher dimensionality does result in more nformation gained. However, the findings should be interpreted with some degree of caution. Firstly, our payment vehicle, a contribution to a fund, might not be as incentive compatible as the costs respondents face in transportation surveys dealing with a private good. Secondly, following largely the modelling approach presented by Caussade et al. (2005) we have not addressed taste heterogeneity so far. As the survey was conducted nation-wide and respondents lived in quite different landscapes, it is likely that participants prefer different changes in their landscapes. Accounting for both observed and unobserved heterogeneity thus is likely to have an impact on the results. Studies investigating the relationship between taste and scale heterogeneity found that they are associated with each other (Dellaert et al. 2012; Czajkowski et al. 2014; Scarpa et al. 2011).

In future analysis more flexible models in terms of taste heterogeneity will be applied. Also, we will investigate to which extent the design dimensions influence the number of times the status quo alternative is chosen. The results presented so far in the literature indicate that the status quo option is more likely to be chosen when the choice design is more complex (Boxall et al. 2009; Zhang and Adamowicz 2011). Additionally, effects of the choice task on choices and WTP estimates will be analysed using various measure of complexity (Dellaert et al. 2012; DeShazo and Fermo 2002; Swait and Adamowicz 2001). This could provide further insights about the impact of the design dimensions on the results of stated CE suggesting whether researchers need to care more about this issue, at least within the ranges investigated in this study. Another important topic is, particularly if a larger number of choice sets is used to gain more information per respondent, whether sequences of choice questions are incentive compatibility. Vossler et al. (2012) show that under certain conditions sequences of binary choice questions are incentive compatible. Whether this finding is valid also for other choice question formats, e.g., with three and more alternatives, or other survey contexts is so far an open question.

Acknowledgments

We would like to thank two anonymous reviewers for their very valuable comments. Also we would like to thank Helen Lauff (Link-Institut) for her advice and patience while implementing our Design of Designs approach in the survey software. Funding for this research was provided by the Federal German Ministry of Education and Research (Fkz. 033L029G; Fkz 01LL0909A) and is gratefully acknowledged.

2.7 References

Arentze T, Borgers A , Timmermans H, DelMistro R (2003) Transport Stated Choice Responses: Effects of Task Complexity, Presentation Format and Literacy. Transportation Research Part E 39: 229–244.

Bech M, Kjaer T, Lauridsen J (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Economics 28(3): 273-286.

Boxall P, Adamowicz W L , Moon A (2009) Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. Aust J Agr Resour Economics 53: 503-519.

Bradley M, Daly A (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21: 167-184.

Brouwer R. Dekker T, Rolfe J, Windle J (2010) Choice Certainty and Consistency in Repeated Choice Experiments. Environ Resource Econ 46: 93-109.

Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.

Carlssona F, Mørkbak M R, Olsen S B (2012) The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice of Modelling 5: 19-37.

Caussade S. Ortúzar J, Rizzi L I, Hensher D A (2005) Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B, 39: 621-640.

Chintakayala P K, Hess S, Rose J M, Wardman M (2010) Effects of Stated Choice Design Dimensions on Model Estimates. In: Hess S, Daly A (ed) Choice Modelling: State-of-the-Art and the State-of Practice. Proceedings from the Inaugural International Choice Modelling Conference, Emerald Press, Bingley.

Czajkowski M. Giergiczny M, Greene W H (2012) Learning and Fatigue Effects Revisited. The Impact of Accounting for Unobservable Preference and Scale Heterogeneity on Perceived

Ordering Effects in Multiple Choice Task Discrete Choice Experiments. forthcoming in Land Economics.

Chung C, Boyer T, Han S (2011) How many choice sets and alternatives are optimal? Consistency in choice experiments. Agribusiness 27: 114-125.

Dellaert B, Brazell J, Louviere J J (1999) The effect of attribute variation on consumer choice consistency. Marketing Letters 10: 139-147.

Dellaert B, Donkers B, van Soest A (2012) Complexity Effects in Choice Experiment–Based Models. Journal of Marketing Research 49: 424-434.

DeShazo J R, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 44: 123-143.

Galesic M. (2006) Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. Journal of Official Statistics 22: 313-328.

Heiner R (1983) The origin of predictable behaviour. The American Economic Review 73: 560-595.

Hensher D A, Stopher P A, Louviere J J (2001) An exploratory analysis of the effect of numbers of choice sets in designed choice experiments: an airline choice application. Journal of Air Transport Management 7: 373-379.

Hensher D A (2004) Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. Journal of Transport Economics and Policy 38: 425-446.

Hensher D A (2006) Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment. Environ Resource Econ 34: 7-44.

Hess S, Hensher D A, Daly A (2012) Not bored yet – Revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: 626–644.

Hoerger M (2010) Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. Cyberpsychology, Behavior and Social Networking 13: 697-700.

Hole A R (2006) Small-sample properties of tests for heteroscedasticity in the conditional logit model. Economics Bulletin 3: 1-14.

Louviere, J. J., Carson, R. T., Burgess, l., Street, D., Marley, A.A.J. (2013) Sequential preference questions factors influencing completion rates and response time using online panel. The Journal of Choice Modelling 8: 19-31

Louviere J J, Islam T, Wasi N, Street D, Burgess, L.B. (2008) Designing Discrete Choice Experiments: Do Optimal Designs Come At A Price? Journal of Consumer Research 35: 360-375

Louviere, J.J., D.A., Hensher, Swait J.D. (2000) Stated Choice Methods. Analysis and Application. Cambgridge.

McNair B J, Bennett J, Hensher D A (2011) A comparison of responses to single and repeated discrete choice questions. Resource and Energy Economics 33: 554-571.

Poe G L, Welsh M P, Champ P A (1997) Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. Land Economics 73: 255-267.

Ratcliffe J, Longworth L (2002) Investigating the structural reliability of a discrete choice experiment within health technology assessment. International Journal of Technology Assessment in Health Care 18: 139-144.

Rolfe J, Bennett J (2009) The impact of offering two versus three alternatives in choice modelling experiments. Ecological Economics 68: 1140-1148.

Rose J M , Hensher D A, Caussade S, Ortúzar J, Jou R-C (2009) Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. Journal of Transport Geography 17: 21-29.

Ryan M, Wordsworth S (2000) Sensitivity of Willingness to Pay Estimates to the Level of Attributes in Discrete Choice Experiments. Scottish Journal of Political Economy 47: 504-524.

Scarpa R, Notaro S, Louviere J, Raffaelli R (2011) Exploring scale effects of best/worst rank ordered choice data to estimate benefits of tourism in Alpine grazing commons. American Journal of Agricultural Economics 93: 813-828.

Scarpa R, Rose J M (2008) Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. The Australian Journal of Agricultural and Resource Economics 52: 253-282.

Swait J D, Adamowicz W L (2001) Choice environment, market complexity and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organizational Behaviour and Human Decision Processes, 86: 141-167.

Vicente P, Reis E (2010) Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys. Social Science Computer Review 28: 251-267.

Vossler C A, Doyon M, Rondeau D (2012) Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics 4: 145-171.

Zhang J, Adamowicz W L (2011) Unravelling the Choice Format Effect: A Context-Dependent Random Utility Model, Land Economics 87: 730-743.

2.8 Appendix 1

Table 1: Attribute levels for designs with two levels

Attribute	narrow		base		wide	
Share of forest (%)	-20	+20	-25	+25	-30	+30
Land conversion (%)	-40	+40	-50	+50	-60	+60
Bio_whole	80	90	70	100	53	117
Bio_agrar	76	89	65	100	49	116
Bio_forest	85	95	80	100	62	118
Bio_resident	70	90	60	100	44	116
Bio_other1	82	93	75	100	58	117
Bio_other2	70	90	60	100	44	116
Bio_other3	76	89	65	100	49	116

Table 2: Attribute levels for designs with three levels

Attribute	narrow			base			wide		
Share of forest (%)	-20	0	+20	-25	0	+25	-30	0	+30
Land conversion (%)	-40	0	+40	-50	0	+50	-60	0	+60
Bio_whole	80	85	90	70	85	100	53	85	117
Bio_agrar	76	82	89	65	80	100	49	82	116
Bio_forest	85	90	95	80	90	100	62	90	118
Bio_resident	70	80	90	60	80	100	44	80	116
Bio_other1	82	87	93	75	85	100	58	87	117
Bio_other2	70	80	90	60	80	100	44	80	116
Bio_other3	76	82	89	65	80	100	49	82	116

Table 3: Attribute levels for designs with 4 levels

Attribute	narrow				base				wide			
Share of forest (%)	-20	-8	+8	+20	-25	-10	+10	+25	-30	-12	+12	+30
Land conversion (%)	-40	-20	+20	+40	-50	-25	+25	+50	-60	-30	+30	+60
Bio_whole	80	83	87	90	70	80	90	100	53	74	95	117
Bio_agrar	76	80	84	89	65	77	88	100	49	71	93	116
Bio_forest	85	88	91	95	80	86	94	100	62	80	98	118
Bio_resident	70	76	83	90	60	73	87	100	44	68	92	116
Bio_other1	82	85	89	93	75	83	91	100	58	77	96	117
Bio_other2	70	76	83	90	60	73	87	100	44	68	92	116
Bio_other3	76	80	84	89	65	77	88	100	49	71	93	116

2.9 Appendix 2

Table 1: Linear regression number of drop-outs and design dimensions

Dimension	Coefficient	|t-value|
Number of choice sets	0.6344	2.52
Number of alternatives	1.2264	0.65
Number of attributes	4.2485	2.59
Number of levels	-2.2349	1.09
Wide level range	5.0965	1.26
Narrow level range	2.1755	0.57
Constant	-14.9163	-0.99

Note: n = 16

Table 2: Heteroskedastic logit with order effects CS_2-CS_{24}

Variable	Coefficient	|t-Val.|	Variable	Coefficient	|t-Val.|
ASCsq	0.1842	3.08	CS_1	0.0000	fixed
Share of forest	0.0065	3.11	CS_2	-0.0243	0.32
Land conversion	-0.0034	3.12	CS_3	0.0662	0.90
Bio_whole	0.0043	3.03	CS_4	0.0561	0.76
Bio_agrar	0.0021	2.89	CS_5	0.1601	2.24
Bio_forest	0.0024	2.77	CS_6	0.1818	2.57
Bio_urban	0.0026	2.69	CS_7	0.1518	1.93
Bio_other1	0.0027	2.83	CS_8	0.1020	1.26
Bio_other2	0.0010	2.22	CS_9	0.2437	3.18
Bio_other3	0.0019	2.58	CS_{10}	0.2700	3.56
Cost	-0.0023	3.15	CS_{11}	0.1905	2.45
Age group 31 to 40 years	-0.0134	0.34	CS_{12}	0.1967	2.53
Age group 41 to 50 years	-0.0781	2.08	CS_{13}	0.1677	1.89
Age group 51 to 60 yeras	0.0629	1.61	CS_{14}	0.2737	3.17
Age group older than 60 years	-0.1647	3.48	CS_{15}	0.1618	1.82
Gender male	-0.1287	4.90	CS_{16}	0.2460	2.85
Middle-school	0.0049	0.08	CS_{17}	0.1549	1.73
High-school	0.1015	1.71	CS_{18}	0.2334	2.72
University degree	0.1489	2.63	CS_{19}	0.2367	2.13
Dummy 4 alternatives	0.0910	2.21	CS_{20}	0.2640	2.41
Dummy 5 alternatives	0.0401	1.14	CS_{21}	0.3799	3.63
Number of attributes	0.1496	2.66	CS_{22}	0.3802	3.70
Number of levels	0.3861	4.04	CS_{23}	0.2846	2.59
Interaction attributes and levels	-0.0784	4.24	CS_{24}	0.1851	1.60
Dummy narrow range	0.2318	6.51			
Dummy wide range	-0.0995	2.65			
Log-likelihood null	-31114.18				
Log-likelihood model	-27375.53				
Observations	90354				

3 Stated and Inferred Attribute Non-Attendance in a Design of Designs Approach

Priska Weller[a], Malte Oehlmann[b], Petr Mariel[c], Jürgen Meyerhoff[b]

Abstract

Attribute non-attendance in stated choice experiments (CE) has gained attention in literature, with some studies finding that not all respondents attend to all attributes. While the current studies show that taking non-attendance into account can significantly influence survey results, it is not yet clear what motivates respondents to ignore or pay less attention to some of the attributes. In the present study, we use 16 different split samples designed according to a Design of Designs plan, varying different aspects of dimensionality, i.e., the number of choice sets, the number of alternatives, or the number of attributes. Firstly, to analyse the relationship between stated attribute non-attendance and the design dimensions we test whether both are significantly associated. Secondly, we estimate equality-constrained latent class models with classes based on pre-defined rules to infer attribute non-attendance and analyse the influence of the design dimensions. Overall, the results indicate a rather weak relationship between stated or inferred attribute non-attendance and design dimensions. However, an interesting finding is that there is an association with the number of alternatives and with the number of sets.

Keywords: attribute non-attendance, design of designs, stated choice experiment, latent class analysis

Published in: Journal of Choice Modelling (2014), 11: 43-56,
https://doi.org/10.1016/j.jocm.2014.04.002

[a] Johann Heinrich von Thünen Institute, Institute for International Forestry and Forest Economics, Leuschnerstraße 91, 21031 Hamburg, Germany

[b] Institute for Landscape and Environmental Planning, Technische Universität Berlin, Straße des 17. Juni 145, 10623 Berlin, Germany

[c] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Avda. Lehendakari Aguirre, 83, 48015 Bilbao, Spain

3.1 Introduction

Recently, the question whether respondents to stated choice experiments (CE) attend to all attributes presented on the choice sets has gained quite a bit of interest. Results from several studies indicate that not all respondents attend to all attributes and that attribute non-attendance (ANA) can significantly bias parameter and subsequent willingness to pay (WTP) estimates (Hensher et al., 2005, Scarpa et al., 2009, Campbell et al., 2011). The awareness that ANA can be an issue when estimating discrete choice models leads to the question of the causes of ANA. One source of ANA might be the complexity of the CE as respondents may employ different answer heuristics when the choice sets become more complex (Scarpa et al., 2009, Hensher et al., 2012). So, respondents may focus only on a subset of attributes.

In the present study we interpret the complexity of the choice task as the design dimensionality of the CE. We systematically vary the number of choice sets, the number of alternatives, the number of attributes and their levels as well as the level range following a Design of Designs (DoD) approach introduced by Hensher (2004) in transportation research. Using 16 split samples, this study aims to investigate the relationship between these five design dimensions and the attendance to attributes. Influences of choice task complexity have been analysed in terms of several issues such as error variance (e.g., Caussade et al., 2005, Meyerhoff et al., 2014) or the number of status quo choices (e.g., Boxall et al., 2009). However, we are not aware of any study systematically relating ANA to the design dimensionality of the CE in random utility models (RUMs). To our knowledge only Hensher (2006) investigated choice task complexity influences on ANA analysing stated non-attendance data. Being aware of the design dimensions that cause ANA might be important for the researcher not only in the estimation stage, but particularly in the design stage of the CE.

The data used in this study comes from a nation-wide online survey on land use changes in Germany. The attributes we considered were: Share of Forest, Land Conversion and different biodiversity attributes which were aggregated and disaggregated across designs according to different landscape types. This allowed us to systematically focus on the effect of the design dimensions. After answering the choice questions we asked respondents to state their attendance to the attributes they were presented with in the CE. In total we collected 1684 interviews resulting in an average of more than 100 respondents per treatment.

We analyse the relationship between ANA and the design dimensionality in two different ways. Firstly, we report the answers to non-attendance questions across the 16 different designs and identify those design dimensions that significantly influence stated ANA. Secondly, we model inferred ANA by specifying an equality-constrained latent class model (ECLCM) for each design. We then derive inferences of dimensionality impacts on ANA by comparing the membership probabilities of different ANA classes across the 16 LC models. We apply this two-step approach separately analysing stated and inferred ANA since it has been argued that even when respondents stated that they have not attended to an attribute, the models indicate that people have at least partially taken them into account (e.g., Hensher and Rose, 2009). Furthermore, both Scarpa et al. (2013) and Kragt (2013) compared two approaches, one not incorporating and one incorporating stated ANA. Kragt (2013) did this by setting attribute parameters to zero where ANA had been stated by respondents. They found that the constrained LC model, which did not include stated ANA, better matched observed data.

The paper is structured as follows: a literature review is presented in Section 2. Section 3 then explains the design of the study and the sampling procedure. Next, Section 4 outlines the hypotheses to be tested in this study and the modelling approach while Section 5 presents results. Section 6 concludes with a discussion of the results and their implications.

3.2 Literature review

The literature on attribute non-attendance in stated choice studies can be divided broadly into two main approaches: the stated non-attendance approach and the analytical non-attendance approach. In the Appendix a table with studies investigating ANA is provided. In the former, respondents are asked directly whether they have considered all attributes describing the alternatives of the choice tasks or whether they have ignored one or more attributes while choosing among them. Stated ANA is usually investigated using a binary response variable with the information on ANA commonly being collected after answering all choice sets (serial stated ANA). Alternatively, non-attendance questions may be asked after each choice set (e.g., Scarpa et al., 2009). The answers to these questions are then used to put certain restrictions on the RUMs. Among the first studies investigating ANA was Hensher et al. (2005). They used the answers to a binary non-attendance question to specify a mixed logit model in which respondents who stated that they ignored a certain attribute were expected to have zero utility. This approach of individual-level zero marginal utility weights has subsequently been applied in several other studies including Hensher (2006), Hensher et al. (2007) and Kragt (2013). Another way to implement the information gained from a stated non-attendance question into RUMs is to estimate different coefficients for each attribute: one for the group of respondents who stated that they did not ignore the attributes and one for those who stated that they had (Hess and Hensher, 2010; Scarpa et al., 2013; Colombo et al., 2013).

However, the reliability of the stated ANA approach has been put into question. Firstly, it has been argued that a respondent may assign low importance to an attribute although stating that he or she ignored it completely (e.g., Hess and Hensher, 2010). This would lead to an overestimation of ANA (Carlsson et al., 2010). Nevertheless, it should be noted that in some instances it was found that respondents who report having ignored an attribute do indeed have zero marginal utility for that attribute (Balcombe et al., 2011). Secondly, directly incorporating the responses to the non-attendance questions into the RUM may cause potential problems of endogeneity bias (Hess and Hensher, 2010). Thirdly, there is also an increase in literature raising the – still unanswered – question of the reliability of complementary questions regarding non-attendance (Hensher and Rose, 2009; Hensher and Greene, 2010).

Given the limitations of the stated non-attendance approach, there has been increasing interest in literature which analyses different methods of inferring attribute non-attendance. These studies infer the information from the data by using diverse econometric treatments. The modelling approach that has probably been implemented most consists of LC models in which a probabilistic decision process captures the attendance to attributes imposing specific restrictions on the utility expressions for each class. The majority of studies using discrete probability distributions implemented the ECLCM, where ANA is operationalised by allowing some respondents to belong to latent classes with zero utility weights for selected attributes, while non-zero parameters are assumed to take the same values across classes (Scarpa et al., 2009; Hensher and Greene, 2010; Campbell et al., 2011; Hess et al., 2013; Scarpa et al., 2013).

Nevertheless, Hess et al. (2013) argue that the latent class approach so far applied might be misguided as the results might be confounded with regular taste heterogeneity. They suggest instead a combined LC mixed logit model that allows jointly for attribute non-attendance and continuous taste heterogeneity. One of their findings is that non-attendance is substantially reduced in these models. Similarly, Hensher et al. (2013) presented different latent class models accounting not only for ANA, but also for aggregation of common-metric attributes. One of their models comprises several full attendance classes while other classes account for ANA regarding certain attributes. The model does not constrain the parameters to be equal across classes. In the next step they add to this model an additional layer of heterogeneity by specifying some parameters to follow a random distribution. For their data they found that accounting additionally for taste heterogeneity through random parameters within a latent class only marginally improves the model, but increases the probability of membership to full attribute attendance classes. Collins et al. (2013) presented a generalised random parameters attribute nonattendance model (RPANA) as an alternative approach; besides, they found that with stated ANA as covariates the model performance was improved.

Studies which employ both the stated non-attendance and the inferred non-attendance approach are Hensher et al. (2007), Hensher and Rose (2009), Campbell et al. (2011), Kragt (2013) and Scarpa et al. (2013). The overall finding is that results from analytical and stated ANA are not consistent, and that the inferred approach provides a better model fit. However, independent of the modelling approach almost all studies find that accounting for ANA improves model fit, indicating that respondents have indeed applied different processing strategies (Hensher et al., 2005; Hess and Hensher, 2010; Scarpa et al., 2013). Moreover, accounting for ANA leads to significantly different willingness to pay estimates which may be lower (Hensher, 2006; Campbell et al., 2008; Scarpa et al., 2009) or higher (Hensher et al., 2005; Hensher and Rose, 2009; Hensher and Greene, 2010; Lagarde, 2013).

Systematic relationships between ANA and the design dimensionality of the CE have so far only been investigated in one study. Hensher (2006) used an ordered heterogeneous logit model with the dependent variable defining the number of attributes that were stated to be ignored, and the independent variables being, among others, the dimensionality. Hensher (2006) found evidence that all design dimensions influence stated ANA. However, we are not aware of any study that relates ANA to the design dimensionality of the CE by either implementing the responses to a stated ANA question into a RUM or inferring ANA from the data.

3.3 Study design and samples

The non-price attributes used in the stated CE are all related to environmental aspects associated with land use changes. The list of attributes (Table 1) comprises Share of Forest, Land Conversion, several attributes regarding biodiversity conservation and a price attribute Cost. Cost is presented as an annual contribution to a landscape fund. All attributes except those concerning biodiversity conservation were presented in all designs. Moreover, for Cost attribute levels were not varied across designs. To adjust the number of attributes according to the design plan, the attribute Biodiversity Conservation was based on an indicator using stocks of bird populations used in Germany (BMU 2010). This indicator can be segregated and aggregated into various bird populations in different landscapes (e.g., birds in the whole landscape can be split up into birds in agrarian landscape plus birds in other landscapes).

Table 1: Attributes used in the CE

Attribute	Description
Share of Forest	Percentage changes in the share of forest
Land Conversion	Percentage changes in land conversion for housing development and traffic
Bio_whole	Biodiversity in the whole landscape including all landscape types
Bio_agrar	Agricultural landscape biodiversity
Bio_forest	Forest landscape biodiversity
Bio_urban	Urban area biodiversity
Bio_other1	Biodiversity in other landscape types: Forests, urban areas, mountains, waters
Bio_other2	Biodiversity in other landscape types: Urban areas, mountains, waters
Bio_other3	Biodiversity in other landscape types: Mountains, waters
Cost	Contribution to a landscape fund in € per year

Following the design master plan by Hensher (2004), 16 different designs were created using NGENE software. The numbers of choice sets used were 6, 12, 18 or 24, and we presented three to five alternatives including a status quo option. The number of attributes ranged from four to seven. The number of attribute levels varied from two to four and the range in attribute levels had three specifications: A base range, a narrow range (base -20%), and a wide level range (base +20%). From these dimensions, C-efficient designs were created. They minimise the variance of the marginal WTP estimates (Scarpa and Rose, 2008). Uniform priors were used, ranging from 0.1 to 1.2 with a positive sign for the non-cost attributes and a negative sign for cost. However, due to time constraints we did not update the design while the survey was conducted. Each respondent was randomly allocated to one of the designs. In Table 2 the different combinations of design dimensions are presented; the design numbers (1-16) indicate the sequence as seen in Hensher (2004).

Table 2: Design overview

Design	Sets	Alternatives	Attributes	Levels	Range
1	24	4	5	3	Base
2	18	4	5	4	Base+20%
3	24	3	6	2	Base+20%
4	12	3	6	4	Base
5	6	3	4	3	Base+20%
6	24	3	4	4	Base-20%
7	6	4	7	2	Base-20%
8	12	5	4	4	Base+20%
9	24	5	4	4	Base
10	6	5	7	3	Base+20%
11	6	4	6	4	Base-20%
12	12	5	5	2	Base-20%
13	18	4	7	2	Base
14	18	3	4	3	Base-20%
15	12	3	5	2	Base
16	18	5	6	3	Base-20%

All choice sets included a status quo alternative, i.e., a zero price option with no environmental changes, plus two or more alternatives depending on the Design of Designs plan. Respondents were requested to choose the alternatives on the choice sets considering land use changes within a radius of about 15 kilometres around their place of residence. At the end of the questionnaire, standard socio-demographic information was requested from respondents. For more details about the study design and sampling see Meyerhoff et al. (2014).

3.4 Hypotheses and modelling approach

As we employ both stated and inferred non-attendance, the same hypotheses are investigated for both approaches. Our general assumption is that not all dimensions equally affect non-attendance. We expect that the more attributes presented on a choice set the more likely it is to observe ANA (Hypothesis 1). The reason is that respondents find it too difficult to attend to all attributes or taste heterogeneity occurs where some of the attributes are not found to be relevant for choices. Similarly, the number of alternatives is expected to influence ANA as well. Increasing the number of alternatives makes CEs larger and probably more complex, thus respondents are more likely to adopt a simplifying answer heuristic (Hypothesis 2). Furthermore, we expect that ANA is positively associated with an increasing number of choice sets (Hypothesis 3). As respondents answer more choice sets they are likely to focus more on those attributes they find important and ignore the others. Regarding the number of attribute levels, we expect that the more different levels are present on a choice set the lower is ANA (Hypothesis 4). The reason is that respondents are more likely to find levels that are relevant to them. Finally, regarding attribute range we have no specific expectation regarding the relationship between either. At first glance, varying the attribute range does not change the appearance of the choice set as an increasing number of attributes or alternatives changes a choice set. Hensher et al. (2012) found that attribute levels and range could trigger ANA, but argue that the reason for their finding might be rather respondent-specific. Furthermore, Hensher (2006) found that ANA depends on the difference between experienced levels of a reference alternative and levels on other choice sets.

The relationships between both stated and inferred ANA and the design dimensions are analysed using different approaches. Firstly, we employ a non-parametric chi-squared test to determine whether there is a relationship between the design dimensions and stated non-attendance (two categorical variables). Secondly, to infer non-attendance we employ an ECLCM with pre-defined classes for attribute non-attendance. In this case we use Spearman's rank correlation coefficient to analyse the relationship between the estimated class membership probabilities for the pre-defined classes and the design dimensions as the class membership probabilities represent a continuous variable. We decided to confine the analyses of ANA to the attributes Share of Forest, Land Conversion and Cost. The reason for this is that these attributes are part of all 16 designs, thus they allow us to draw conclusions about all possible relationships. Due to the split-up according to the design plan, each of the biodiversity attributes appears only in a subset of designs.

3.4.1 Investigating stated attribute non-attendance

After the sequence of choice tasks, we gathered information concerning the degree of stated attendance from answers to a categorical five-point scale. This scale comprised the following answer options: 'never', 'rarely', 'sometimes', 'often' and 'always'. For each attribute, respondents were requested to state their degree of attendance[1] (see Scarpa et al., 2013, for a similar approach). Using this information we calculate overall stated attendance to all attributes across designs, including a graphical representation for each design dimension. Furthermore, we analyse a possible relationship between the design dimensions and stated ANA to the attributes selected for the analysis by the use of a non-parametric chi-squared test. For this we recode answers given by the respondents into binary categories. The new first category 'not attended

1 Translation of the question: 'If you think back to your choice decisions, to what extent did you attend to the individual attributes of the various alternatives in your choices overall?'

to' represents the original answer options 'never', 'rarely' and 'sometimes' and the new second category 'attended to' represents the original answer options 'often' and 'always'. For example, the chi-squared test then investigates the significance of the relationship between the four levels of the design dimension 'number of choice sets' (6, 12, 18, 24 choice sets) and two answer categories (non-attendance and attendance). Additionally, we generate scatter plots with frequencies of responses for each relationship. From these we analyse whether a change in any of the design dimensions decreases or increases the frequency of attending to an attribute.

3.4.2 Investigating inferred attribute non-attendance

To investigate the relationship between inferred non-attendance and the design dimensions we use ECLCM with pre-defined classes. The basic idea of these models is to estimate the probability of belonging to one of the pre-defined classes[2] for each individual. In each of these classes, selected attributes carry zero utility weights for reflecting non-attendance. Additionally, the attribute coefficients to be estimated have been constrained to be equal across classes. Therefore, these models do not account for preference heterogeneity across respondents, i.e., the coefficient estimates are the same for each attribute across classes.

The following five classes are used in each model, i.e., for each of the 16 designs:

- For one class we assume that all attributes have been attended to: this is the full attendance class.
- We define a separate class for non-attendance to each of the attributes Share of Forest, Land Conversion, and Cost, by setting the coefficient of that attribute to zero, implying ANA.
- For the fifth class we assume that none of these three attributes has been attended to, thus all coefficients are set to zero; this is the random choice class.

---

2 During the model selection process we tested various approaches. We particularly tried to account for taste heterogeneity as recent studies indicate that not accounting for taste heterogeneity may overestimate attribute non-attendance (Hess et al., 2013; Hensher et al., 2013). However, neither approach was successful. Even if it worked for one of the designs it was not possible to apply the approach successfully across all 16 designs, i.e., 16 independent samples. Firstly, an increasing number of so called 'full-attendance classes' were added to the equality-constrained non-attendance classes as was done, for example, by Hensher et al. (2013). The information criteria (BIC and CAIC) indicated in the majority of designs that adding more classes would be beneficial while at the same time model output could be interpreted as being less and less meaningful. For example, in many classes the cost coefficient had a positive sign and membership probabilities became very small. Secondly, we tried to add another layer of heterogeneity in the latent class model by specifying some attributes to follow a random distribution in the full attendance class. Although we tried many different model setups, these models never converged except for single data sets out of our 16 samples. Thus, neither approach was useable to infer non-attendance using the same approach across all 16 designs.

The results from the ECLCM[3] give us the probabilities that all attributes have been attended to, one of the three investigated attributes has not been attended to, or none of the attributes has been attended to. However, in the model we did not account for response strategies ignoring any two attributes jointly. We then analyse a possible relationship between design dimensions and class probabilities by means of Spearman's rank correlation coefficient. Additionally, we again generate scatter plots associating the probabilities of belonging to a certain class and the design dimensions. This way we graphically analyse whether a positive or negative relationship between a design dimension and the share of respondents who did not attend to an attribute in a specific design is present.

3.5 Results

3.5.1 Socio-demographic description

The Germany-wide online survey was run in December 2012. Respondents were recruited from a panel of a survey company. Each respondent was randomly allocated to one of the 16 designs. The number of respondents ranged from 79 to 222 as we previously agreed with the survey company to interview different numbers of respondents per design to account for the different numbers of choice observations resulting from each design. In total, 1684 fully completed interviews were conducted. The average interview length was 23 minutes and the response rate[4] was 29.5%. Table 3 reports the number of interviews realised and basic socio-demographics for all 16 designs. The relative shares of education refer to the highest educational level achieved.

3 Due to limited space, utility parameters generated in the models can be obtained from the authors upon request.

4 The response rate was calculated by dividing the number of interviews completed by the number of persons invited for an interview.

Table 3: Socio-demographics across designs

Design	Interviews Completed	Mean age in years	Males in %	Less than middle-school degree in %	Middle-school degree in %	High-school degree in %	University degree in %
1	85	41.98 (14.41)	44.71 (0.50)	4.71 (0.21)	29.41 (0.46)	22.35 (0.42)	43.53 (0.50)
2	82	44.32 (14.31)	41.46 (0.50)	6.10 (0.24)	24.39 (0.43)	20.73 (0.41)	48.78 (0.50)
3	124	43.81 (16.60)	42.74 (0.50)	7.26 (0.26)	25.00 (0.43)	28.23 (0.45)	39.52 (0.49)
4	80	42.50 (14.25)	50.00 (0.50)	6.25 (0.24)	32.50 (0.47)	30.00 (0.46)	31.25 (0.47)
5	104	43.30 (14.59)	54.81 (0.50)	4.81 (0.21)	24.04 (0.43)	25.00 (0.44)	46.15 (0.50)
6	82	42.32 (13.21)	46.34 (0.50)	7.32 (0.26)	29.27 (0.46)	18.29 (0.39)	45.12 (0.50)
7	222	42.49 (13.12)	46.85 (0.50.)	5.41 (0.23)	31.08 (0.46)	27.48 (0.45)	36.04 (0.48)
8	79	42.54 (12.75)	56.96 (0.50)	10.13 (0.30)	25.32 (0.44)	29.11 (0.46)	35.44 (0.48)
9	81	43.47 (14.06)	37.04 (0.50)	4.94 (0.22)	28.40 (0.45)	30.86 (0.48)	35.80 (0.48)
10	151	42.02 (13.69)	49.67 (0.50)	10.60 (0.31)	20.53 (0.41)	29.80 (0.46)	39.07 (0.49)
11	88	41.44 (13.81)	56.62 (0.50)	4.55 (0.21)	36.36 (0,48)	21.59 (0.41)	37.50 (0.49)
12	81	39.47 (13.29)	39.51 (0.50)	3.70 (0.19)	22.22 (0.42)	38.27 (0.49)	35.80 (0.48)
13	113	42.09 (13.15)	55.75 (0.50)	4.42 (0.21)	25.66 (0.44)	26.55 (0.44)	43.36 (0.50)
14	84	43.40 (12.42)	0.50 (0.50)	15.48 (0.36)	23.81 (0.44)	23.81 (0.43)	36.90 (0.49)
15	144	40.44 (13.60)	0.50 (0.50)	7.64 (0.27)	29.86 (0.46)	26.39 (0.44)	36.11 (0.48)
16	84	42.15 (13.80)	48.81 (0.50)	5.95 (0.24)	23.81 (0.43)	28.57 (0.45)	41.67 (0.50)
Total	1684	42.33 (13.59)	48.34 (0.50)	6.83 (0.25)	27.08 (0.44)	26.84 (0.44)	39.25 (0.49)

Note: Standard deviations in parentheses.

3.5.2 Overall stated attribute non-attendance across designs

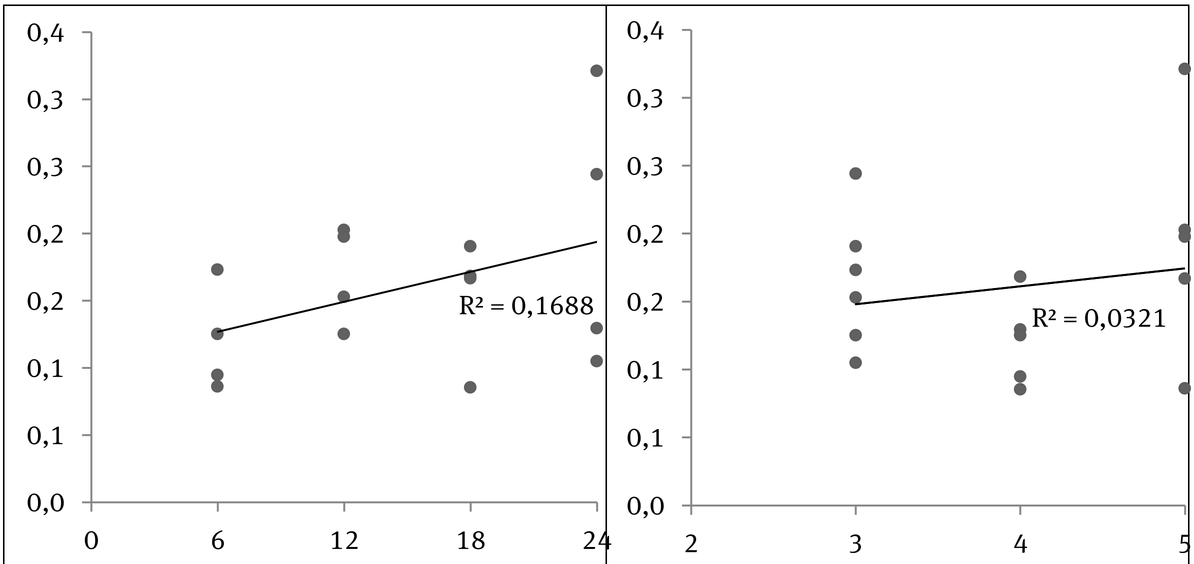
As a first step in our analysis, we calculated the share of respondents who stated that they 'often' or 'always' attended to all attributes presented in the choice sets. Table 4 presents these results in descending order of stated attribute attendance per design.

Table 4: Stated attendance to all attributes (percentage of respondents who answered 'often' or 'always' attended to)

Design	9	6	8	12	14	5	13	16	15	1	4	11	3	7	10	2
Stated Att.	32	24	20	20	19	17	17	17	15	13	13	13	10	9	9	9

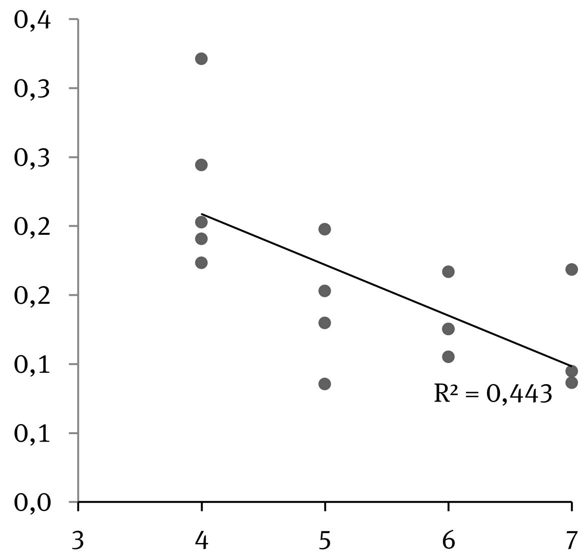
Figure 1 presents the results from Table 4 separately for the five design dimensions: number of choice sets, number of alternatives, number of attributes, number of levels and level range in ascending order of the levels of each design dimension. A linear trend line is added each time. As can be seen, only in graph c of Fig. 1 a clear negative relationship between stated non-attendance and the design dimension is visible: attendance is higher the fewer attributes a design comprises. This is in line with our first hypothesis, that a lower number of attributes is associated with higher stated attribute attendance. The other four plots do not show clear positive or negative relationships. The preliminary graphical analysis leads, therefore, to the conclusion that the number of choice sets, the number of alternatives, the number of levels and the level range does not have an influence on the stated attendance to attributes.

Figure 1: Stated attribute attendance per design dimension

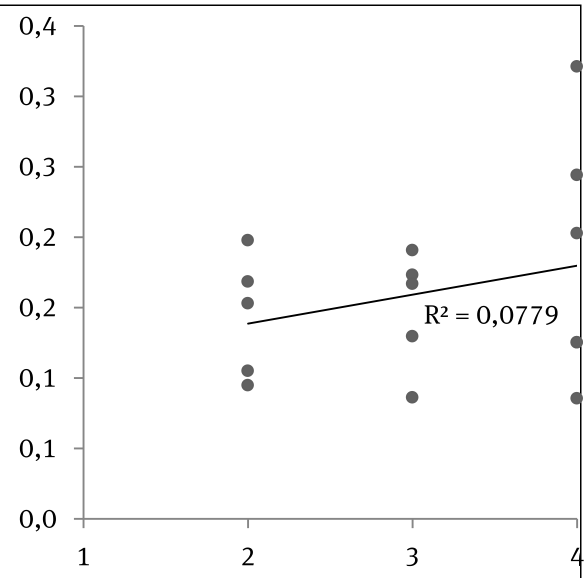


1a) Number of choice sets

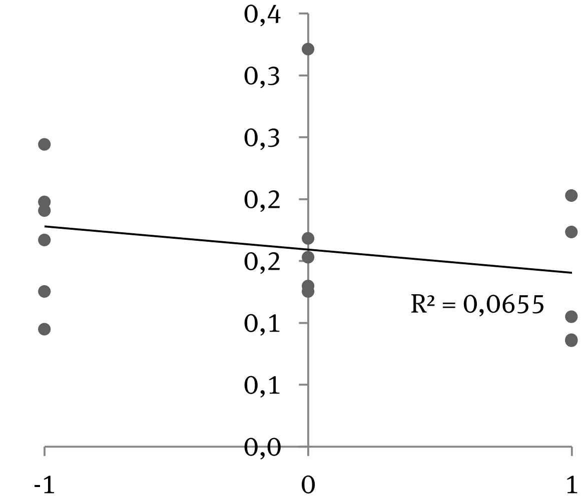
1b) Number of alternatives



1c) Number of attributes



1d) Number of levels



1e) Level range

Legend:
X-axis – levels of design dimension
Y-axis – stated attendance (in decimal numbers)
Grey dots – Stated attendance
Black lines – Linear (Stated attendance)

3.5.3 Attribute specific stated non-attendance across designs

The responses to the non-attendance scale are reported in Table 5. The mean values, calculated from response scores for each attribute separately, indicate the degree of ANA to each single attribute in each design. Values closer to five indicate higher levels of attendance while values closer to one indicate higher levels of non-attendance.

The attribute Share of Forest receives rather high stated attendance values (almost always > 4 and mean 4.10, meaning it was 'often attended to') whereas the stated attendance values for other attributes are relatively lower (almost always < 4). Attendance to Land Conversion in some designs also lies around 4, but mean attribute attendance over all designs is 3.75. Attendance to the Cost attribute is 3.29, meaning it was 'sometimes attended to'; this effect is visible in all designs. This finding is in line with other literature. Campbell et al. (2008) found in their study that Cost was taken into account by 69% of respondents, less than all other attributes presented.

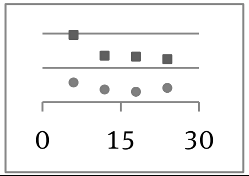
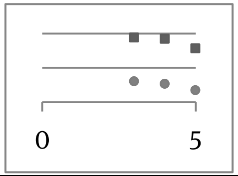
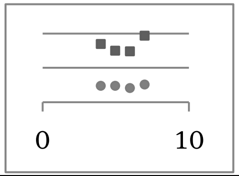
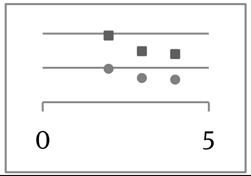
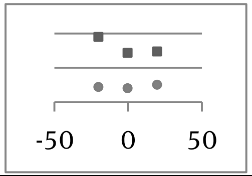
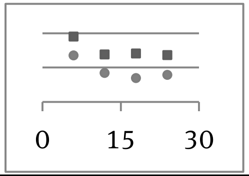
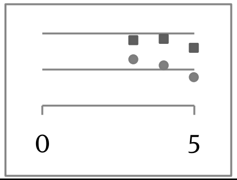
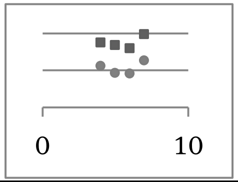
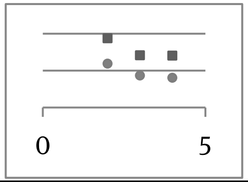
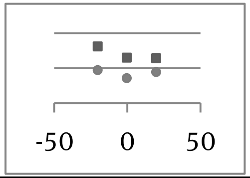
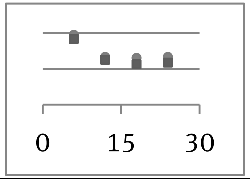
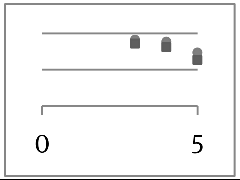
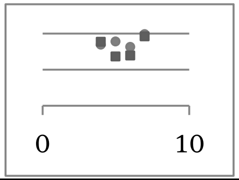
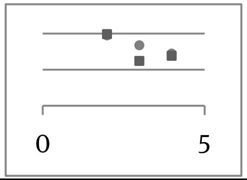
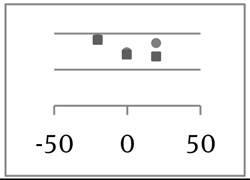
Table 5: Means of attribute attendance values coded on a five-point scale

Attributes	Design																Mean
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
Share of Forest	4.16	3.99	3.87	4.03	3.90	3.91	4.13	4.01	4.26	4.15	4.20	4.35	4.20	4.22	3.93	4.36	4.10
Land Conversion	3.82	3.75	3.65	3.64	3.50	3.52	3.65	3.58	4.01	3.73	3.76	3.97	3.89	3.85	3.64	4.05	3.75
Cost	3.06	3.01	3.11	3.23	3.53	3.50	3.38	3.22	3.45	3.03	3.38	3.47	3.52	3.23	3.25	3.20	3.29

Note: Number of observations for Share of Forest, Land Conversion and Cost: N = 1684.

Next, we use a chi-squared test to investigate the relationship between two categorical variables, the five design dimensions and the two recoded stated answer categories to single attributes across designs (non-attendance and attendance). From scatter plots we aim to deduce the direction of the influence of the design dimensions on stated attendance. Figure 2 presents *p*-values from the chi-squared tests (significant correlations are bold) and corresponding scatter plots. As the results reveal, for the majority of attributes we cannot find significant or clear relationships between stated attendance and the design dimensions. Nevertheless, for some combinations of attributes and design dimensions this relationship is significant. A significant negative effect on stated attendance to Land Conversion was detected for the number of choice sets presented. This means that the more choice sets are presented the less likely it becomes that a respondent states that he/she has attended to the attribute. The number of alternatives has a significant and overall negative influence on stated attendance to Share of Forest and Land Conversion. For the number of attributes no significant relation was detected. For the number of levels included in the designs, we observe a negative and significant effect on stated attendance to Cost. Interestingly, this effect occurs despite the fact that for the Cost attribute neither the number nor the range of levels (i.e., the price) varies. A significant negative effect was also observed for the influence of the level range on Share of Forest. If the level range becomes wider, less respondents answer that they attended to Share of Forest.

Figure 2: Chi-squared tests and scatter plots

Attribute	Number of Sets	Number of Alternatives	Number of Attributes	Number of Levels	Level Range
Share of Forest	*p-value=0.218*  	*p-value=0.040*  	*p-value=0.762*  	*p-value=0.584*  	*p-value=0.018* 
Land Conversion	*p-value=0.055*  	*p-value=0.017*  	*p-value=0.646*  	*p-value=0.669*  	*p-value=0.153* 
Cost	*p-value=0.886*  	*p-value=0.743*  	*p-value=0.118*  	*p-value=0.020*  	*p-value=0.130* 

Note: Correlations significant at the 10% level or higher are bold; the units of the y-axis are not presented to have a larger scatter plot area. X-axis: levels of design dimension; Y-axis: frequencies of responses on five-point scale in binary coding; black square: ‘attended to’, grey dot: ‘not attended to'.

3.5.4 Inferred attribute non-attendance across designs

Table 6 reports class membership probabilities of ANA inferred from the ECLCM. The last column reports the mean, minimum and maximum probabilities across all designs. The attribute with the highest average probability of ANA is the Cost attribute (average 27.2%). The maximum predicted class size for non-attendance to Cost is 40.1% (Design 10), while the minimum predicted class size is 10.8% (Design 3). Adding the probability of not attending to any attribute (class 'Random choice') implies that approximately 41.9% of respondents may not have attended to the Cost attribute. Next, the attribute Share of Forest has a mean class probability of ANA of 18.1% (class 'Share of Forest not attended to'). It varies from 37.4% (Design 3) to 8.5% (Design 16). Again, adding the probability of not attending to any attribute results in a probability of ANA to Share of Forest of 32.8%. The third non-attendance class investigated here, non-attendance to Land Conversion, has a mean class probability of 17.0% across designs with the highest probability equal to 29.4% (Design 11) and the lowest probability equal to 8.2% (Design 10). The overall probability after adding the mean membership probability for the class 'Random choice' results in a probability of 31.7%. Finally, across all designs respondents have an average probability of attending to all attributes of 23.1%, ranging from 17.1% (Design 12) to 33.0% (Design 14).

Next, we turn to the relationship between the design dimensions of the 16 samples and the probabilities to attend to or not attend to attributes according to the pre-defined classes.

Table 6: Class membership probabilities in %

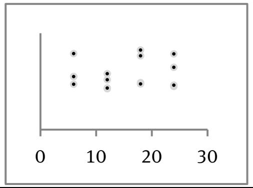
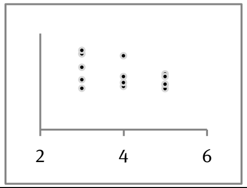
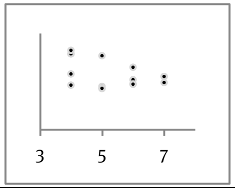
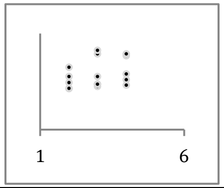
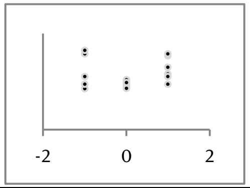
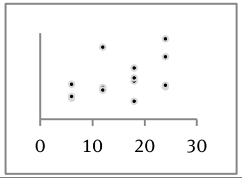
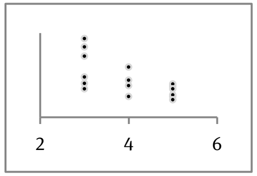
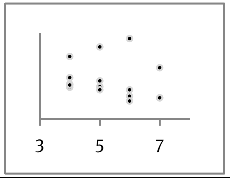
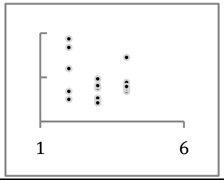
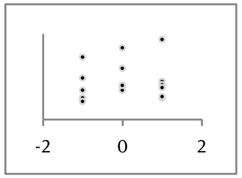
Attendance class	Design 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean Min Max
All attended to	17.9 (4.2)	30.7 (5.3)	25.9 (4.0)	20.8 (5.5)	31.5 (5.9)	31.5 (5.3)	22.1 (4.1)	23.2 (5.1)	18.5 (4.3)	22.1 (4.1)	18.9 (4.9)	17.1 (4.3)	19.5 (4.0)	33.0 (5.3)	17.2 (3.5)	19.0 (4.3)	23.1 17.1 33.0
Share of Forest not attended to	15.1 (3.9)	17.7 (4.4)	37.4 (4.4)	13.6 (4.5)	16.2 (4.5)	29.0 (5.1)	10.0 (4.0)	14.9 (4.4)	15.9 (4.1)	10.0 (4.0)	10.7 (4.1)	13.7 (3.8)	23.9 (4.0)	19.3 (4.3)	33.5 (4.1)	8.5 (3.0)	18.1 8.5 37.4
Land Conversion not attended to	19.7 (4.5)	11.8 (3.8)	20.8 (3.8)	25.6 (5.4)	9.1 (3.8)	15.0 (4.1)	8.2 (4.1)	16.2 (4.7)	18.5 (4.3)	8.2 (4.1)	29.4 (6.0)	13.9 (4.3)	21.7 (4.1)	15.3 (4.0)	18.5 (3.6)	19.5 (4.7)	17.0 8.2 29.4
Cost not attended to	28.5 (5.1)	29.5 (5.3)	10.8 (3.0)	27.4 (5.6)	30.3 (5.7)	14.5 (4.2)	40.1 (6.2)	21.9 (4.7)	33.4 (5.2)	40.1 (6.2)	24.7 (5.8)	39.9 (5.7)	25.1 (4.4)	17.7 (4.5)	17.4 (3.6)	33.2 (5.3)	27.2 10.8 40.1
Random choice	18.8 (4.5)	10.3 (3.5)	5.1 (2.2)	12.6 (4.5)	12.9 (4.6)	9.9 (3.7)	19.6 (2.9)	23.9 (5.1)	13.7 (3.8)	19.6 (2.9)	16.4 (5.5)	15.4 (4.4)	9.8 (2.9)	14.7 (4.0)	13.3 (3.1)	19.8 (4.8)	14.7 5.1 23.9

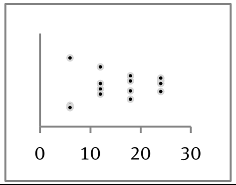
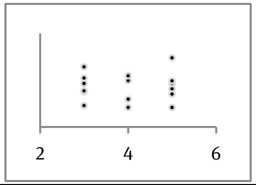
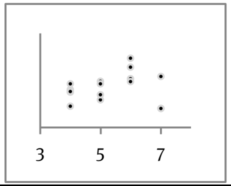
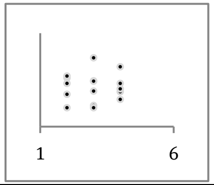
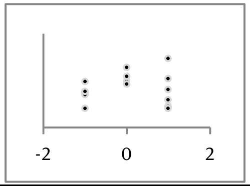
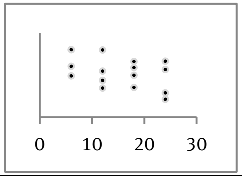
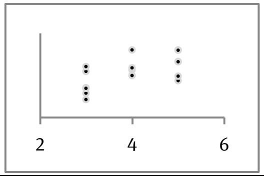
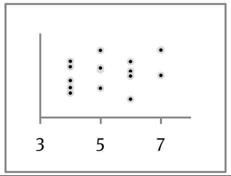
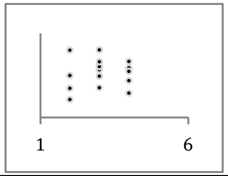
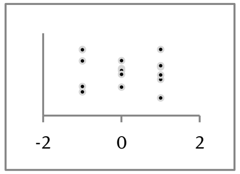
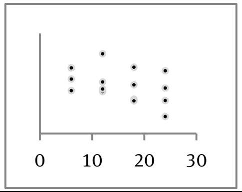
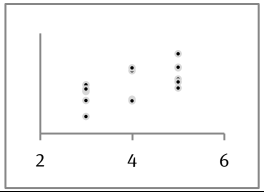
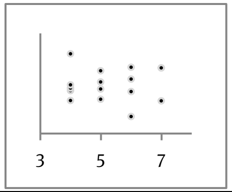
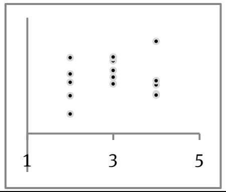
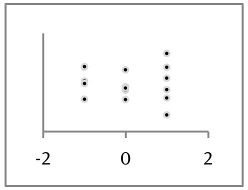
Note: Standard errors in parenthesis.

Spearman's rank correlation coefficient was used to analyse the relationship between single design dimensions and class membership probabilities. Figure 3 presents the correlation coefficients together with the associated *p*-values and scatter plots for each combination of analysed design dimensions and classes included in the ECLCM. These scatter plots should help to infer the direction of a possible relationship, i.e., whether an increasing design dimension affects inferred ANA positively or negatively.

It becomes obvious from Figure 3 that only one of the design dimensions has a systematic influence on inferred attribute non-attendance. The number of alternatives in four out of five cases significantly influences the probabilities of attending to or not-attending to an attribute. The probability to attend to all attributes decreases with an increasing number of alternatives, which is in line with our expectations that with an increasing dimensionality the probability of ANA also increases. Interestingly, the probability to attend to the attribute Share of Forest increases with an increasing number of alternatives. For the Cost attribute we see that with an increasing number of alternatives the probability of not attending to Cost increases. Also, and also in line with our expectations, the probability of random choice increases when choice sets comprise more alternatives. Furthermore, the only relationship that is statistically significant is between the number of sets and the non-attendance to the attribute Share of Forest. According to the scatter plot this relationship is positive, which is in line with our expectation that an increasing number of choice sets positively influences the probability of not attending to this attribute. For the other relationships we could not find significant associations. Thus, overall the results indicate that with an increasing number of alternatives respondents seem to focus on the two attributes Share of Forest and Land Conversion, while paying less attention to the other attributes, especially the Cost attribute.

Figure 3: Spearman's rank correlation coefficient

Attribute	**Number of Sets**	**Number of Alternatives**	**Number of Attributes**	**Number of Levels**	**Level Range**
All attended to	*rho=0.03* *p-value=0.91* 	***rho= -0.48*** ***p-value=0.06*** 	*rho= -0.27* *p-value=0.32* 	*rho=0.33* *p-value=0.22* 	*rho=0.17* *p-value=0.53* 
Share of Forest not attended to	***rho=0.53*** ***p-value=0.03*** 	***rho= -0.67*** ***p-value=0.00*** 	*rho= -0.40* *p-value=0.13* 	*rho= -0.11* *p-value=0.68* 	*rho=0.10* *p-value=0.71* 

Land Conversion not attended to	*rho=0.33* *p-value=0.22* 	*rho= -0.04* *p-value=0.89* 	*rho=0.19* *p-value=0.49* 	*rho= -0.01* *p-value=0.97* 	*rho=0.09* *p-value=0.73* 
Cost not attended to	*rho= -0.40* *p-value=0.12* 	***rho=0.57*** ***p-value=0.02*** 	*rho=0.30* *p-value=0.25* 	*rho= -0.04* *p-value=0.87* 	*rho= -0.12* *p-value=0.66* 
Random choice	*rho= -0.41* *p-value=0.11* 	***rho=0.67*** ***p-value=0.00*** 	*rho=0.06* *p-value=0.82* 	*rho=0.04* *p-value=0.87* 	*rho= -0.12* *p-value=0.67* 

Note: Correlations significant at the 5% level or higher are bold; the units of the y-axis are not presented to have a larger scatter plot area. X-axis: levels of design dimension; Y-axis: class membership probabilities from Table 6.

3.6 Discussion

The aim of this study was to investigate the relation between the dimensionality of a CE and ANA. To determine the influence of these dimensions we applied a Design of Designs approach resulting in 16 different treatments. Stated non-attendance was recorded using a five-point non-attendance response scale while for inferring non-attendance we used an ECLCM.

The results from the analysis of stated ANA show that the attribute Share of Forest was favoured by many respondents, i.e., they attended to it to a large extent across all designs ('often attended to'). Stated attendance to the Cost attribute was rather low ('sometimes attended to'). Mean attendance to Land Conversion lay between the two.

We investigated the relation between stated attendance to single attributes and design dimensions from chi-squared tests, which revealed only five statistically significant relations. Scatter plots derived from frequencies of responses showed that higher numbers of choice sets tend to negatively influence the frequency of stating 'attendance to' an attribute (e.g., to the attribute Land Conversion). A higher number of alternatives tends to negatively influence attendance to attributes (e.g., to the attribute Share of Forest or Land Conversion). Also, a higher number of levels negatively influenced the probability of stating 'attendance to' the Cost attribute, also a wider range of levels negatively influenced stated attendance (e.g., for Share of Forest).

Inferred ANA was derived from an ECLCM, which estimates class membership probabilities for pre-defined classes ('All attended to', 'Share of Forest not attended to', 'Land Conversion not attended to', 'Cost not attended to', 'Random choice/No attribute attended to'). These showed that attendance differs quite substantially between the classes. Particularly worth mentioning is that – according to our latent class model – only a relatively small proportion of respondents considered all attributes. The class size for this group of respondents varied between 17 and 33%. On the other hand, the probabilities for random choice varied between 5 and 24% indicating that respondents allocated to some designs have a significant probability to not consider any attribute.

Using Spearman's rank correlation coefficient we deduced that higher numbers of choice sets positively influenced the probability of non-attending to Share of Forest. Higher numbers of alternatives significantly influenced four out of five class membership probabilities investigated: probabilities of 'all attributes attended to' and 'Share of Forest not attended to' classes decreased with more alternatives, but 'Cost not attended to' and 'Random choice' classes increased with more alternatives.

Regarding attendance to Cost, we found that a large share of the respondents scarcely attended to Cost even though attention to the Cost attribute is crucial from an economic point of view to calculate marginal willingness to pay and welfare measures. This result is visible in both stated and inferred attendance: Respondents stated that it was 'sometimes' included in the decision and class membership probabilities for 'Cost not attended to' lay between 11 and 40%, meaning that a substantial number of respondents did not include the Cost attribute in their decision. Similar results were found in other studies (e.g., see Campbell et al., 2008, for inferred results; Hensher et al., 2012, Kragt, 2013). Reasons for ANA to Cost could be that as long as costs lie within a certain range, they are not considered important. We know this from focus group discussions where participants indicated that they don't care about costs unless they are beyond a certain threshold. For designs with a high number of attributes, ANA to Cost may additionally be driven by more important attributes being offered, which attract more attention than the less important Cost attribute. We believe that social desirability may also have affected stated non-attendance to Cost. It may be that respondents wanted to conceal their true cost awareness when asked about attribute attendance, thus they stated less attendance to Cost than actually present.

Looking at our results, the first, and probably most important, conclusion is that in both approaches – stated and inferred – the relationship between attribute attendance and design dimensions is not very strong. Nevertheless, when going into detail, our findings regarding stated and inferred ANA suggest some relationships between the design dimensions and ANA. The results from stated ANA indicate that a higher number of alternatives are significantly influencing this form of attendance, and the results from inferred ANA also indicate that the number of alternatives is the main driver behind non-attendance. Regarding our hypotheses the analysis found no significant influence of the number of attributes on ANA, which does not confirm our first hypothesis that the number of attributes is a driver of non-attendance. However, the second hypothesis, stating that an increasing number of alternatives influences ANA, has been supported by the results from the stated and the inferred approach. Also, the expected direction of the effect for an increase in the number of alternatives has been found for the majority of classes (fewer 'all attended to', more 'Cost not attended to', more 'Random choice'). An increasing number of alternatives could make comparisons, and subsequently the choosing of an alternative, more difficult. In the end this might lead to ignoring attributes as a response to the higher level of difficulty. The results of the two approaches are consistent with the third hypothesis, stating that an increasing number of choice sets may increase non-attendance. From the stated ANA analysis we deduced that Land Conversion was attended to less when more choice sets were presented. From the inferred analysis we deduced that more choice sets led to more ANA to Share of Forest. It may be that respondents perceived a higher number of sets as exhausting and just focused on some attributes. The fourth hypothesis, stating that more levels lead to more attendance, was not supported by the stated analysis. Here, more levels led to less attendance to the Cost attribute.

Asking respondents to state their level of attendance has been a subject of debate, i.e., some researchers question whether respondents correctly state to what extent they have or have not attended to certain attributes. However, from the literature it is not clear whether models inferring ANA really do better. The ECLCMs in particular are suspected to overestimate ANA because they do not account for taste heterogeneity (Hess et al., 2013, Hensher et al., 2013). Therefore, our models might overestimate the degree of ANA and should be interpreted with a degree of caution. On the other hand, as outlined in Section 4, we estimated various models incorporating taste heterogeneity via additional full attendance classes or additional random parameters within classes, but they were not informative or sufficiently interpretable. This is especially true as we tried to apply a consistent approach across all 16 designs, i.e., samples. It may have been possible to adjust individual models to each sample, this would, however, have required a very high effort. Hensher et al. (2013) report in a footnote that a huge amount of time was necessary to identify what they thought of as an appropriate model. To do this for all 16 samples would have been beyond our resources and based on our experience with these models, we think that it is questionable whether the insights gained would have justified this effort.

To conclude, from what we found in this study, results indicate that the connection between dimensionality and non-attendance to attributes is not very strong. This is true for both stated and inferred ANA. Nevertheless, based on our results, we suggest that a negative effect on attribute attendance should be expected from the number of alternatives and the number of choice sets. In contrast, we would not expect that a higher number of attributes, levels and level range results in higher degrees of ANA in future CE. Future research options in this field, apart from other model specifications or the application of the RPANA model (Collins et al., 2013), could be to consider other attribute processing rules such as the aggregation of common-metric attributes. For this analysis, the biodiversity attributes could be reconsidered as this decision rule might be used for these attributes.

Acknowledgements

We would like to thank two anonymous reviewers for their very helpful comments. Also, we would like to thank the editor, David Hensher, for his support. The authors acknowledge financial support from the Basque Government through grant IT-642-13 and from UPV/EHU through grant US12/09. Furthermore, funding for this research was provided by the German Federal Ministry of Education and Research (Fkz. 033L029G; Fkz 01LL0909A) and is gratefully acknowledged.

3.7 References

Balcombe, K., M. Burton and D. Rigby, 2011. Skew and attribute non-attendance within the Bayesian mixed logit model. *Journal of Environmental Economics and Management*, 62, 446-461.

Boxall, P., W. L. Adamowicz and A. Moon, 2009. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *The Australian Journal of Agricultural and Resource Economics*, 53, 503-519.

Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.

Campbell, D., D. A. Hensher and R. Scarpa, 2011. Non-attendance to attributes in environmental choice analysis: a latent class specification. *Journal of Environmental Planning and Management*, 54, 8, 1061-1076.

Campbell, D., W. G. Hutchinson and R. Scarpa, 2008. Incorporating Discontinuous Preferences into the Analysis of Discrete Choice Experiments. *Environmental and Resource Economics*, 41, 3, 401-417.

Carlsson, F., M. Kataria and E. Lampi, 2010. Dealing with Ignored Attributes in Choice Experiments on Valuation of Sweden’s Environmental Quality Objectives. *Environmental and Resource Economics*, 47, 65-89.

Caussade, S., J. de Dios Ortuzar, L. I. Rizzi and D. A. Hensher, 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*, 39, 621-640.

Collins, A. T., J. M. Rose and D. A. Hensher, 2013. Specification issues in a generalised random parameters attribute nonattendance model. *Transportation Research Part B*, 56, 234-253.

Colombo, S., M. Christie and N. Hanley, 2013. What are the consequences of ignoring attributes in choice experiments? Implications for ecosystem service valuation. *Ecological Economics*, 96, 25-35.

Hensher, D. A., 2004. Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. *Journal of Transport Economics and Policy*, 38, 3, 425-446.

Hensher, D. A., 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, 21, 861-878.

Hensher, D. A., A. T. Collins and W. H. Greene, 2013. Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: a warning on potential confounding. *Transportation*, 40, 5, 1003-1020.

Hensher, D. A. and W. H. Greene, 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical Economics*, 39, 2, 413-426.

Hensher, D. A. and J. M. Rose, 2009. Simplifying choice through attribute preservation or non-attendance: implications for willingness to pay. *Transportation Research Part E*, 45, 583-590.

Hensher, D. A., J. M. Rose and T. Bertoia, 2007. The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies. *Transportation Research Part E*, 43, 73-89.

Hensher, D. A., J. M. Rose and W. H. Greene, 2005. The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes. *Transportation*, 32, 203-222.

Hensher, D. A., J. M. Rose, and W. H. Greene, 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation*, 39, 235-245.

Hess, S. and D. A. Hensher, 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B*, 44, 6, 781-790.

Hess, S., A. Stathopoulos, D. Campbell, V. O'Neill and S. Caussade, 2013. It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation*, 40, 3, 583-607.

Hole, A. R., 2011. A discrete choice model with endogenous attribute attendance. *Economics Letters*, 110, 203-205.

Hole, A. R., J. R. Kolstad and D. Gyrd-Hansen, 2013. Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behaviour & Organization*, 96, 21-31.

Kragt, M. E., 2013. Stated and Inferred Attribute Attendance Models: A Comparison with Environmental Choice Experiments. *Journal of Agricultural Economics*, 64, 3, 719-736.

Lagarde, M., 2013. Investigating Attribute Non-Attendance and its Consequences in Choice Experiments with Latent Class Models. *Health Economics*, 22, 554-567.

Meyerhoff, J., M. Oehlmann and P. Weller, 2014. The Influence of Design Dimensions on Stated Choices in an Environmental Context. *Environmental and Resource Economics*, (accepted).

Rose, J. M., D. A. Hensher, W. Greene and S. Washington, 2012. Attribute exclusion strategies in airline choice: accounting for exogenous information on decision maker processing strategies in models of discrete choice. *Transportmetrica*, 8, 5, 344-360.

Scarpa, R., T. J. Gilbride, D. Campbell and D. A. Hensher, 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, 36, 2, 151-174.

Scarpa, R. and J.M., Rose, 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *The Australian Journal of Agricultural and Resource Economics*, 52, 253-282.

Scarpa, R., M. Thiene and D. A. Hensher, 2010. Monitoring Choice Task Attribute Attendance in Nonmarket Valuation of Multiple Park Management Services: Does It Matter? *Land Economics*, 86, 4, 817-839.

Scarpa, R., R. Zanoli, V. Bruschi and S. Naspetti, 2013. Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, 95, 1, 165-180.

3.8 Appendix

Literature regarding attribute non-attendance

Study	**Stated ANA**		**Inferred ANA**		**Key findings**
	Model	Treatment in RUM	Model	Treatment in RUM	
Hensher et al. (2005)	Mixed Logit	Individual-level 0-utility weights			Higher WTP when not accounting for ANA
Hensher (2006)	Mixed Logit	Individual-level 0-utility weights			Lower WTP when accounting for ANA
Hensher et al. (2007)	Mixed Logit	Individual-level 0-utility weights, expected maximum utility			Higher WTP when accounting for ANA
Campbell et al. (2008)	MECLM	Individual-level 0-utility weights, parameterisation of the scale parameter based on attribute processing strategies			Better model fit, lower WTP and different scale parameters when accounting for ANA
Hensher and Rose (2009)	Mixed Logit	Individual-level 0-utility weights			Better model fit and higher WTP when accounting for ANA
Scarpa et al. (2009)			ECLCM; STAS	0-utility weights (a priori); 0-weight (from data)	Better model fit and lower WTP when accounting for ANA
Carlsson et al. (2010)	Mixed Logit	Individual-level 0-utility weights			No significant differences in preferences and WTP in samples accounting and not accounting for stated ANA
Hensher and Greene (2010)			ECLCM	0-utility weights	Higher WTP when accounting for ANA
Hess and Hensher (2010)	MNL; Mixed Logit	Separate parameters for attendance and non-attendance groups		Coefficient of variation for the conditional distribution	Results from inferred and stated ANA are not consistent, better fit for the inferred approach
Scarpa et al. (2010)	MNL	Individual-level and choice set specific (choice-set ANA) 0-utility weights			Better model fit and more efficient WTP estimates when accounting for choice-set ANA

Balcombe et al. (2011)	Mixed Logit	Unrestricted and restricted interactions, individual-level 0-utility weights			Better model fit when accounting for ANA
Campbell et al. (2011)			Scale-adjusted ECLCM	0-utility weights	Lower WTP when accounting for ANA
Hensher et al. (2012)			ECLCM		ANA and WTP also depend on relevance of levels and range to respondents
Rose et al. (2012)	Mixed Logit	Ignored attributes excluded from estimation procedure			Significant differences in parameter estimates when accounting for ANA
Collins et al. (2013)			(Generalised) RPANA		Generalised RPANA links stated and inferred ANA, better performance than existing models
Colombo et al. (2013)	Mixed Logit	Separate parameters for attendance and non-attendance groups			Accounting for 'sometimes attendance' offers advantages over only 2 attendance groups
Hensher et al. (2013)			Random parameter LCM	0-parameters (mean and SD) + class heterogeneity	Accounting for ANA gives more model fit than random parameters, random parameters increase probability of full-attendance
Hess et al. (2013)			ECLCM; LCM; LC-MMNL	Coefficients set to zero / estimated (point value, continuous distribution)	Better model fit in combined LC-MMNL, implied rate of ANA reduced
Hole et al. (2013)	EAAModel; MEAA	Indicator variable		Attribute not-attended to constrained to 0	Better fit but similar WTP when accounting for ANA
Kragt (2013)	XML; Mixed Logit	Individual-level 0-utility weights	ECLCM	0-utility weights	Best fit for the LCM, little concordance between stated and inferred ANA
Lagarde (2013)			ECLCM	0-utility weights	Better model fit and higher WTP when

					accounting for ANA, no difference for predicted probabilities
Scarpa et al. (2013)	MXL; MNL	Separate parameters for attendance and non-attendance groups	ECLCM	0-utility weights	Best fit for the ECLCM

Note: MECLM = Multinomial error component logit model; (M)EAAModel = (Mixed) endogenous attribute attendance model (Hole, 2011); STAS = Stochastic attribute selection; MXL = Random parameter model with error components; LC-MMNL = Latent class mixed logit model

4 Uncovering Context-induced Status Quo Effects in Choice Experiments

Malte Oehlmann[a], Jürgen Meyerhoff[a], Petr Mariel[b], Priska Weller[c]

Abstract

In this study five design dimensions are varied systematically investigating context-induced status quo effects in choice experiments. Additionally, two structural complexity measures, entropy and the number of attribute level changes, are used to capture status quo effects from the similarity between alternatives and the number of trade-offs. A crucial finding is that the frequency of status quo choices is negatively associated with the number of alternatives indicating preference matching effects. By contrast, the probability of choosing the status quo increases with a higher number of choice tasks, a wider level range, and the similarity between alternatives. Status quo choices are further affected by the current environmental situation perceived by respondents. We also find that marginal and non-marginal welfare estimates are significantly affected by the choice design. One key finding is that the most used choice task format in environmental economics, i.e., two hypothetical and a status quo alternative, is likely to increase the propensity to choose the status quo option.

Keywords: Choice Experiment; Status Quo Effect; Choice Context; Complexity; Design Dimensions; Entropy; Error Component Mixed Logit

Published in: Journal of Environmental Economics and Management (2017), 81: 59–73
http://dx.doi.org/10.1016/j.jeem.2016.09.002

[a] Technische Universität Berlin, Institute for Landscape Architecture and Environmental Planning, Straße des 17. Juni 145, D-10623 Berlin, Germany

[b] Department of Applied Economics III (Econometrics and Statistics), University of the Basque Country (UPV/EHU), Avda. Lehendakari Aguirre, 83, 48015 Bilbao, Spain

[c] Johann Heinrich von Thünen Institute, Institute of Forest Economics, Leuschnerstraße 91, D-21031 Hamburg, Germany

4.1 Introduction

When individuals make choices among alternatives, they have been observed to disproportionally often remain at the status quo option. Samuelsson and Zeckhauser (1988) found this phenomenon to be present in a series of decision-making experiments and termed it the “status quo bias” or “status quo effect”. In subsequent research individuals’ propensity to stay at the status quo has been investigated and explained in several studies within disciplines such as economics, psychology or decision theory (Hartman et al., 1991, Kahneman et al., 1991, Tversky and Shafir, 1992, Dhar, 1997). Findings of this research indicate that status quo effects may manifest themselves due to multiple causes such as loss aversion (Kahneman et al., 1991), the omission bias (Ritov and Baron, 1994), preference uncertainty (Dhar, 1997) or complexity issues (Beshears et al., 2008, Dhar, 1997). They have also been observed in choice experiments (Adamowicz et al., 1998, Scarpa et al., 2005, Hess et al., 2014).

Investigating the status quo effect is particularly important in environmental economics since a large number of choice experiments (CE) conducted in this area do offer, along with alternatives that contain improvements or prevent deteriorations, the opportunity to stay with the present situation. In transportation research, for example, it is often reasonable to use forced-choices as people have to commute to their working place. Here, staying at home and not choosing any of the offered alternatives would not be an option. Concerning environmental changes, however, in the majority of cases it is plausible not to select an alternative with management actions and a positive price, but to “leave the market without buying anything” (Bennett and Adamowicz, 2001). In line with this, the standard CE design, applied in the majority of studies in environmental valuation, includes a status quo alternative and two hypothetical options. The crucial question is then, however, to what extent choices of the alternative with a zero-price and the current environmental situation are due to preferences, i.e., respondents do not care about the good in question or agree with the current situation, or whether choices of the status quo alternative are induced by the choice context.

Generally, we follow Scarpa et al. (2005) and define status quo effects as “the systematic inclination of respondents to display a different attitude towards status quo alternatives from those reserved to alternatives involving some change, over and beyond what can be captured by the variation of attributes’ levels across alternatives”. With respect to the choice context we follow Zhang and Adamowicz (2011), considering the characteristics of the choice format or the choice environment, which are described by a set of context attributes. In the literature several context attributes have been associated with status quo choices. Boxall et al. (2009), for instance, found a positive relationship between the number of choice tasks in the sequence and the tendency to stay at the current situation. The opposite effect was observed when the number of alternatives was increased from two to three or from four to five, respectively (Adamowicz et al., 2011, Rolfe and Windle, 2012). The probability of status quo choices has also been observed to decrease as the number of trade-offs increases (Rolfe and Bennett, 2009, Zhang and Adamowicz, 2011). In contrast, a positive effect on this probability results from alternatives being more similar to each other (Balcombe and Fraser, 2010, Zhang and Adamowicz, 2011). Additionally, the literature indicates that also perceptions of the current situation influence the likelihood to choose the status quo alternative (Marsh et al., 2011). This might indicate to what extent the probability of choosing the status quo alternative is influenced by the quality of the environment respondents perceive in their surroundings. People who perceive the quality of the environment in their surroundings as higher might be less likely to choose alternatives that describe improvements compared to the status quo alternative.

To investigate context-induced status quo choices we apply a Design of Designs (DoD) approach originally introduced by Hensher (2004) (see also Caussade et al., 2005). To our knowledge, effects of the design dimensions have until today in the environmental field only been investigated by varying selected design dimensions, e.g., two versus three alternatives (Rolfe and Bennett, 2009, Zhang and Adamowicz, 2011). Implementing 16 different split samples we systematically vary the number of choice tasks, the number of alternatives, the number of attributes and their levels as well as the level range. In addition to the design dimensions we employ three structural measures of choice task complexity which are derived from the configuration of information conveyed by the choice tasks (DeShazo and Fermo, 2002). Entropy is used to measure the similarity between alternatives in a choice task (Shannon, 1948, Swait and Adamowicz, 2001), and cumulative entropy, accordingly, reflects the accumulating burden induced by the similarity between alternatives when a respondent moves through the sequence of choice tasks. Additionally, the number of attribute level changes in the task is applied as a proxy for the number of trade-offs to be made by respondents (DeShazo and Fermo, 2002). The data is from a web survey dealing with land use changes in Germany.

For our analysis we simultaneously adopt two approaches previously presented in the literature. Firstly, we include the alternative specific constant of the status quo (ASCsq) alternative in the deterministic part of the utility function and interact sets of covariates that describe the choice context as defined in this study with the ASCsq. The covariates are grouped in sets capturing i) complexity impacts from the dimensionality of the different designs, ii) structural measures of choice task complexity, and iii) individual’s perceptions of the current environmental quality as stated by respondents prior to the CE. All models also include interactions among the ASCsq and socio-demographic characteristics of the respondents. Secondly, we introduce different error structures within the stochastic component of the utility function to capture differences in the substitution patterns between the status quo and the non-status quo alternatives (Scarpa et al., 2005, Hess and Rose, 2009, Willis, 2009). In order to investigate whether the choice context affects welfare measures we compare marginal willingness to pay (WTP) estimates and compensating variation measures across those designs that share the same number of attributes.

The contribution of this research is to simultaneously account for the influence of multiple context variables on the status quo effect. In particular, to our knowledge this is the first time that status quo choices are investigated using a DoD approach systematically varying the five design dimensions. Besides improving the general understanding of the relationship between choice context and status quo choices, the findings of this study are expected to have two-fold implications. Firstly, being aware of the factors that may promote the status quo effect is of importance for both researchers and practitioners because the dimensionality and thus the complexity are determined within the design stage of the CE. Secondly, knowing the utility associated with the status quo alternative is crucial for obtaining unbiased WTP and, particularly, non-marginal welfare estimates (Adamowicz et al., 1998, Scarpa et al., 2005, Von Haefen et al., 2005, Adamowicz et al., 2011).

We find the probability of choosing the status quo to be positively related to the number of choice tasks, the similarity between alternatives as well as the level range. In contrast, it is negatively associated with the number of alternatives. The propensity to stay at the status quo is further affected by participants' perception of the current situation. Highly significant error components for alternatives different from the status quo reveal the presence of a stochastic status quo effect. Taking the example of designs with the lowest and highest number of status quo choices, marginal WTP and, particularly, non-marginal welfare measures are found to differ significantly.

4.2 Study design and survey

4.2.1 Dimensionality of choice tasks

Following the design master plan by Hensher (2004), 16 different designs were created. Accordingly, the following dimensions of a CE are varied: number of choice tasks, number of alternatives, number of attributes, number of attribute levels, and the level range. The dimensions of each design are reported in Table 5. The number of choice tasks presented to respondents varies, for example, from 6 to 24, and the number of attributes varies between four and seven. The non-cost attributes are all related to environmental aspects associated with land use changes in Germany. The list of attributes comprises Share of Forest, Land Conversion as well as several attributes regarding biodiversity conservation. In addition, "annual contribution to a landscape fund" was implemented as a payment vehicle.

All attributes except those concerning biodiversity conservation were presented in all designs. The attribute Biodiversity is based on an indicator using stocks of bird populations in Germany (BMU 2010). In order to expand the number of attributes according to the design plan, this indicator can be segregated and aggregated into various bird populations in different landscapes (e.g., birds in the whole landscape can be split up into birds in agricultural landscapes plus birds in other landscape types). Table 1 gives an overview of the attributes used.

Table 1: Attributes used in the CE

Attribute	Label	Level variation
Relative share of forest in the landscape	*Share of Forest*	Percentage changes (positive and negative)
Rate of converting land for infrastructure, urban areas, etc.	*Land Conversion*	Percentage changes (positive and negative)
Biodiversity in the whole landscape including all landscape types	*Bio_whole*	Indicator values
Agricultural landscape biodiversity	*Bio_agrar*	Indicator values
Forest landscape biodiversity	*Bio_forest*	Indicator values
Urban area biodiversity	*Bio_urban*	Indicator values
Biodiversity in other landscape types: Forests, urban areas, mountains, waters	*Bio_other1*	Indicator values
Biodiversity in other landscape types: Urban areas, mountains, waters	*Bio_other2*	Indicator values
Biodiversity in other landscape types: Mountains, waters	*Bio_other3*	Indicator values
Contribution to a landscape fund	*Cost*	€ per year

On each choice task respondents were asked to choose their preferred option among two to four generic, hypothetical alternatives considering landscape changes within a radius of about 15 kilometres (km) around respondents' place of residence and a status quo alternative. The status quo alternative was described and coded as follows: The levels of the first two attributes, Share of Forest and Land Conversion, were presented to respondents as "as today" and coded zero. As we do not know the current state of these attributes for individual respondents, no absolute values could be stated. For the biodiversity attributes we informed respondents about the current state of the indicator for each attribute (e.g. “today 65 on index”) in each choice task. We further informed people that the indicator is an average population measure and might be subject to regional variations throughout Germany. The Cost attribute was presented and coded as “0 €“ in the status quo alternative. This alternative was always shown in the last column of the choice task and had the same levels across all treatments. As payment vehicle we used a contribution to a fund which might not be as incentive compatible as compulsory payments such as taxes. The reason for this was that respondents were asked to value rather local changes within a 15 km surrounding of their place of residence. Tax increases for such a change in land use would not have been plausible for the respondents due to the German tax system. However, in the preamble of the CE it was emphasized that respondents had to pay for the land use changes and that the amount to pay would only be used to implement the policy.

In order to assign the attribute levels to alternatives we created for each of the 16 designs separate Bayesian efficient designs with uniform priors for the attribute parameters to allocate the attribute levels to alternatives. The value ranges for the priors were derived based on previous projects and published studies. The design was generated for a multinomial logit model[5] using the C-error as a design criterion. C-efficient designs aim at minimizing the sum of the variance of the marginal WTP estimates (Scarpa and Rose 2008). To allow for uncertainty in the value of the priors, modified latin-hypercube sampling (ChoiceMetrics 2012) was applied and 1000 draws were taken for each parameter prior from uniform distributions. The algorithm stopped searching for a more efficient design when no improvements with respect to the optimisation criteria were found within 30 minutes. Designs were not blocked and thus a respondent received each time the complete design helping to avoid potential confounding with any response heterogeneity that might be a consequence of blocking (see Cherchi and Hensher 2015). Before respondents could answer the choice tasks, which were presented in a randomized order, they were asked to evaluate the current situation regarding the attributes presented to them in the CE. This was done to obtain information on the perceived status quo which was recorded on a four-point scale ranging from “very low” to “very high”.

5 Although all designs were optimized for a multinomial logit model, we use in our analysis mainly the mixed logit model. At the time when we designed the survey we assumed that designs generated based on the MNL model would also perform - relatively - efficient when panel RPL models are estimated (Bliemer and Rose, 2011) and thus were not concerned that this procedure would result in major losses of efficiency nor that it would impact on the estimates. However, as evidence is becoming available that the neutrality assumption regarding the role of the experimental design might not be justified (e.g., Yao et al., 2015; Fosgerau and Börjesson, 2015), we agree with an anonymous reviewer that our procedure could have implications and thus think that the relationship between experimental design and model outcomes need further research in the near future.

4.2.2 Survey administration and sample

The Germany-wide online survey was run in December 2012. Respondents were recruited from a panel of a survey company. Each respondent was randomly allocated to one of the 16 designs. Since the statistical efficiency differs across designs, we defined targets for some designs with respect to the number of interviews. As a guideline for defining targets we used the information provided by Ngene software (ChoiceMetrics 2012) on the minimum number of observations needed for obtaining statistically significant parameter estimates (t-value equal or above 1.96). In total 1661 respondents completed the survey and provided useable interviews. The average interview length was 23 minutes and the response rate was 29.5%. Table 2 reports the number of useable interviews as well as the socio-demographic variables for each design. Particularly with respect to gender, where the figures vary between 36.25% and 56.96%, there seems to be some variation in the means and standard deviations across designs. However, since we focus mainly on methodical issues and control for possible influences of the socio-demographic characteristics on status quo choices, we do not worry too much about the representativeness of each sample.

Table 2: Completed interviews and socio-demographic variables per design

Design	Completed interviews	Mean age in years (SD)	Male respondents %	Mean years of schooling (SD)
1	83	41.43 (14.14)	44.57	14.07 (3.65)
2	81	44.05 (14.19)	41.04	14.42 (3.66)
3	123	43.66 (13.56)	42.27	13.97 (3.53)
4	80	42.50 (14.25)	50.00	13.34 (3.43)
5	102	42.98 (14.49)	54.90	14.40 (3.55)
6	80	42.34 (13.33)	46.25	14.09 (3.74)
7	217	42.31 (13.01)	47.00	13.69 (3.51)
8	79	42.54 (12.75)	56.96	13.61 (3.54)
9	80	43.31 (14.08)	36.25	13.75 (3.47)
10	150	41.98 (13.72)	50.00	13.95 (3.52)
11	88	41.44 (13.81)	56.62	13.60 (3.63)
12	79	39.01 (13.06)	40.50	13.97 (3.25)
13	111	41.79 (13.04)	55.85	14.22 (3.54)
14	81	43.22 (12.43)	50.61	13.54 (3.70)
15	144	40.44 (13.60)	50.00	13.60 (3.56)
16	83	41.83 (13.56)	49.40	14.08 (3.51)
Total	1661	42.15 (13.53)	48.40	13.88 (3.54)

Note: Standard deviations (SD) in parentheses.

4.3 Modeling approach

4.3.1 Structural complexity measures

In addition to the design dimensions of a CE we use three measures of choice task complexity capturing effects of the structure and configuration of information in a CE. The first measure we use is entropy. Originally defined in information theory, it is a measure of information content or uncertainty (Shannon 1948) and was introduced into the discrete choice literature by Swait and Adamowicz (2001). It can be seen as a measure of the similarity between alternatives or the cognitive burden of the choice task. Entropy is calculated using

$$H(\pi_x) = -\sum_{j=1}^{J} \pi(x_j) \log \pi(x_j) \tag{1}$$

where alternatives of the CE are characterized by $\{x_j, j = 1, \dots, J\}$, and their probability distribution is $\pi(x)$. This way the number of alternatives presented is directly included in the formula and influences the level of entropy. Thus, the more equal the probabilities of the single alternatives, the higher the entropy (Swait and Adamowicz 2001). In this study we calculate entropy using the logit formula with attribute levels recoded to vary between -1 and 1. Therefore, they are orthogonal to each other. This ensures that no attribute level dominates expected utility (Zhang and Adamowicz 2011). Moreover, we use flat priors having equal weights for each attribute (Swait and Adamowicz 2001). Secondly, cumulative entropy measures the uncertainty across all choice tasks or the cumulative cognitive burden of choosing the best option (Swait and Adamowicz 2001). It is calculated by using

$$H_r = \sum_{r'=1}^{r-1} H'_r \text{ for } r = 2, \dots, R \qquad H_r = 0 \text{ for } r = 1 \tag{2}$$

where H_r stands for the cumulative entropy experienced by the respondent up to the r^{th} choice task. Thirdly, the number of attribute level changes occurring across all alternatives in a choice task indicates how many trade-offs respondents face when choosing among the alternatives on a task. More trade-offs between more different attribute levels might increase the cognitive burden of trading-off alternatives against each other (DeShazo and Fermo 2002; Zhang and Adamowicz 2011).

4.3.2 Econometric approach

As point of departure for our analysis we use McFadden's (1974) random utility model. Under this framework, the utility U_{njt} for respondent n of alternative j (from a total of J_n) in choice situation t (from a total of T_n choice occasions) is given by:

$$U_{njt} = V_{njt} + \varepsilon_{njt} = \lambda V^*_{njt} + \lambda \varepsilon^*_{njt}, \quad \text{with} \quad V_{njt} = \sum_{k=1}^{K} \beta_{nk} x_{njtk} + ASC_j, \tag{3}$$

for $n = 1,2, \dots, N, t = 1,2, \dots T_{n,} j = 1,2, \dots, J_n$, where V_{njt} depends linearly on observable explanatory variables, which are usually attributes x_{njtk} and attribute parameters β_{nk}. These parameters are constant ($\beta_{nk} = \beta_k, \forall n$) in classical multinomial logit models (MNL) or can vary randomly over individuals in mixed logit family models. In our case, the constant ASC_j is included only in the status quo alternative ($j = 1$), that is why $ASC_j = 0$ for $j = 2,3, \dots, J_n$. The parameter λ is a scale factor inversely proportional to the common standard deviation σ_ε of the Gumbel error terms ε^*_{njt} and it is usually fixed to $\lambda = \pi/(\sqrt{6}\,\sigma_\varepsilon)$ due to the identification so that $Var(\varepsilon_{njt}) = \pi^2/6$. If we separate the cost ($cost_{njt} = x_{njt1}$) and non-cost attributes ($x_{njtk}, k = 2,3, \dots, K$), the utility from equation (3) can be rewritten as

$$U_{njt} = -\beta_{n1} cost_{njt} + \sum_{k=2}^{K} \beta_{nk} x_{njtk} + ASC_j + \varepsilon_{njt}, \tag{4}$$

and equation (4) is called model in preference space. We use models in preference space to investigate the effect of the design dimensions on the propensity to choose the status quo option. To test, however, whether the marginal WTP and the compensating variation measures of certain designs differ we use utility in WTP space models. Their advantage is that they provide more reasonable distributions of WTP avoiding extremely large WTP's (Train and Week 2005). This is important as we specify the cost attribute as following a lognormal distribution in those models capturing unobserved taste heterogeneity. Using the definition that the marginal willingness to pay for an attribute is the ratio of the attribute's coefficient to the cost coefficient $w_{nk} = \beta_{nk}/\beta_{n1}$, utility can be rewritten as

$$U_{njt} = \beta_{n1}(-cost_{njt} + \sum_{k=2}^{K} w_{nk} x_{njtk} + ASC^*{}_j) + \varepsilon_{njt}, \tag{5}$$

$$U_{njt} = \beta_{n1}\left(-cost_{njt} + \sum_{k=2}^{K} w_{nk} x_{njtk} + ASC^*{}_j\right) + \varepsilon_{njt},$$

The utility expressions (4) and (5) are equivalent and any distribution of β_{n1} and β_{nk} in (4) implies a distribution of β_{n1} and w_{nk} in (5) and vice-versa.

As our analysis is focused on status quo choices, we employ a mixed logit (MXL) model and add error components to the utilities of non-status quo alternatives to achieve flexible patterns of substitution via an induced correlation structure across utilities (Scarpa et al. 2005; Scarpa et al. 2007). In the presence of status quo effects different correlation patterns arise between the unobservable components of utility of the status quo alternative and those in non-status quo alternatives. Equation (6) incorporates, therefore, these error components and equation (4) can be rewritten as

$$U_{n1t} = -\beta_{n1} cost_{n1t} + \sum_{k=2}^{K} \beta_{nk} x_{n1tk} + ASC_1 + \varepsilon_{n1t},$$

$$U_{njt} = -\beta_{n1} cost_{njt} + \sum_{k=2}^{K} \beta_{nk} x_{njtk} + u_{nj} + \varepsilon_{njt}, \qquad j = 2,3, \dots, J_n, \tag{6}$$

where $u_{nj} \sim N(0, \sigma^2)$ are additional error components to ε_{njt} which are Gumbel-distributed. ASC_1 labels the status quo alternative and is subsequently indicated by ASCsq.

Now let $j_{n,t}$ be the alternative chosen by consumer n in choice situation t, such that $P_{n,t}(j_{n,t})$ gives the logit probability of the observed choice for consumer n in choice situation t. The logit probability of consumer n's observed sequence of choices is $P_n = \prod_{t=1}^{T_n} P_{n,t}(j_{n,t})$. We assume that β_n is distributed over consumers with density $g(\beta)$. In this case, if $P_{n,t}^{R}(j_{n,t})$ gives the logit probability of the observed choice for consumer n in choice situation t, the logit probability of consumer n's observed sequence of choices is $P_n^R = \int_\beta \prod_{t=1}^{T_n} P_n^R(j_{n,t})\, g(\beta) d\beta$ and the log-likelihood function for the observed choices that is maximised by estimation procedure is $LL = \sum_{n=1}^{N} ln\,(P_n^R)$.

4.3.3 Estimated models

Table 3 presents the main models we estimate in this study. The base model (Model 1) is a plain MNL comprising the choice attributes, the constant term ASCsq and its interactions with socio-demographic variables. Both the ASCsq and these interactions are incorporated in all subsequent models. With Model 2 we estimate an MXL model accounting for unobserved heterogeneity. All attribute parameters are specified as random parameters following a normal distribution except the cost coefficient that is assumed to be lognormally distributed with negative sign. As we assume that respondents across Germany could value the proposed land use changes positively or negatively, the normal distribution for the non-cost attributes is justified. Additionally, this model, as all following MXL models, contains error components capturing the unobserved heterogeneity among the hypothetical alternatives as defined by (6). The next two models extend Model 2 by variables related to design dimensionality of the choice tasks. Model 3 additionally includes interactions between the ASCsq and the five design dimensions. Moreover, Model 3 includes the structural complexity measure entropy as well as the perceived status quo regarding the attributes Share of forest, Land conversion, and Biodiversity. Due to the fact that cumulative entropy and the number of level changes are highly correlated with two design dimensions (see Table 6), only entropy is incorporated as a structural complexity measure. This will be detailed later on. Model 3 is, therefore, defined by equations (6), where socio-demographic variables, entropy, perceived status quo variables together with variables capturing the design dimensions have been added to the utility of the status quo alternative.

As the variation in all design dimensions could cause a potential scale effect, it is assumed that the scale λ in (3) can vary over individuals in the last model (Model 4). That is why, similar to, for example, Caussade et al. (2005) or Zhang and Adamowicz (2011), we specify the scale factor as a function of the six design dimensions ($DD_{kn}, k = 1,2,\ldots,6$) according to $\lambda_n = \exp(\sum_{k=1}^{6} \gamma_k DD_{kn})$. The exponential form ensures a positive scale factor. As in this model the design dimensions are included in scale, they are dropped from the utility function because, as reported by other authors (e.g., Zhang and Adamowicz 2011), we experienced identification problems when incorporating them in both the utility and the scale function. Thus, Model 3 and Model 4 differ only with respect to whether the design dimensions are incorporated in the utility or scale function. Therefore, equations (6) of Model 3 are generalized in Model 4 as follows

$$U_{n1t}^* = (\exp(\textstyle\sum_{k=1}^{6} \gamma_k DD_{kn}))\,(-\beta_{n1} cost_{n1t} + \sum_{k=2}^{K} \beta_{nk} x_{n1tk} + ASC_1) + \varepsilon_{n1t},$$

$$U_{njt}^* = (\exp(\textstyle\sum_{k=1}^{6} \gamma_k DD_{kn}))\,(-\beta_{n1} cost_{njt} + \sum_{k=2}^{K} \beta_{nk} x_{njtk} + u_{njt}) + \varepsilon_{njt}, \quad j = 2,3,\ldots,J_n. \tag{7}$$

In this case, only socio-demographic variables, entropy and perceived status quo variables has been added to the utility of the status quo alternative as the variables considering the design dimensions are included in the scale function. All models were estimated in PythonBiogeme (Bierlaire 2003, 2008), the MXL by maximum simulated likelihood with 2000 Halton draws. In addition to the log-likelihood values we use the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model comparisons.

Table 3: Model specifications

Model	Type	Utility function					Scale	Error
		Main effects	Interactions with ASC of status quo alternative					Component
1	MNL	Attributes	socio-dem.					
2	MXL	Attributes	socio-dem.					error comp.
3	MXL	Attributes	socio-dem.	design dim.	entropy	perceived SQ		error comp.
4	MXL	Attributes	socio-dem.		entropy	perceived SQ	design dim.	error comp.

4.4 Results

4.4.1 Perceived status quo in respondents' surroundings

Table 4 depicts the perceived status quo in the 15 km surroundings of respondents' place of residence with respect to the given share of forests, the rate of land conversion, and the state of biodiversity. The perceived status quo was recorded before respondents faced the sequence of choice tasks.[6] Among the participants 67% stated that the share of forests is "low" or "very low" while the corresponding figure for the state of biodiversity is 60%. For the degree of land conversion, we find that 46% of the participants selected the category "high" or "very high". Remarkably, only 6% of the respondents perceive biodiversity to be very low. The same share of respondents perceives land conversion to be very high.

Table 4: Perceived status quo regarding land use attributes in respondents surroundings (in %)

	Share of Forest	Land Conversion	Biodiversity
Very low	14.39	12.94	5.88
Low	53.10	41.24	54.61
High	27.69	39.80	36.42
Very high	4.86	6.02	3.19
Total	100	100	100

Note: Observations =1661

4.4.2 Status quo choices

We start our analysis by presenting the share of status quo choices per design together with the design dimensions and the structural complexity measures. The first three columns in Table 5 report the design number, the number of respondents, and the number of choices per design. Results are ordered according to the share of status quo choices by design, which is presented in the next column. The share ranges from 16% (Design 16) to 55% (Design 3). This range covers to a large extent the shares of status quo choices reported in the literature (e.g., von Haefen et al. 2005; Hanley et al. 2006; Bullock 2008; Boxall et al. 2009; Lanz and Provins 2015). A glance at the columns reporting the design dimensions reveals a clear link between the share of status quo choices and the number of alternatives on a choice task. The more alternatives a choice task

6 The (translated) wording of this question is: „Considering the 15 km surroundings of your place of residence with respect to the following aspects, how would you describe the current situation?" Subsequently, the aspects, i.e. share of forests, share of land conversion, and biodiversity were briefly described and respondents were asked to describe the present situation in their surroundings on a four-point scale ranging from 'very low' to 'very high'.

comprises, the lower is the share of status quo choices. While the three designs that entail the lowest share of status quo choices all have five alternatives (four designed and the status quo alternative), the six designs with the highest share of status quo choices all have three alternatives (two designed and the status quo alternative). The second, although less clear, link is between the share of status quo choices and the number of attributes. Designs with a higher number of attributes tend to result in a lower share of status quo choices.

The structural complexity measures are reported in the last three columns and do not show a clear association with the share of status quo choices. Generally, mean entropy per choice task varies between 0.39 (Design 15) and 0.98 (Design 9). The design with the lowest share of status quo choices (Design 16) has a rather high value for mean entropy and, due to the large number of choice tasks, a relatively high cumulative entropy value. In contrast, in Design 3, which resulted in the largest share of status quo choices, both mean entropy value and mean number of level changes per choice task are lower than in Design 16. Regarding the number of respondents who have always chosen the status quo alternative, the association with the design dimensions is less obvious. The overall tendency, however, that more people always select the status quo alternative when the number of alternatives decreases remains.

Table 5: Status quo choices, dimensionality and structural complexity measures

Design	**N**	**Choices**	**SQ choices %**	**Always status quo choices %**	**Design dimensions**					**Complexity measures**				
					Number of					Mean entropy per choice task (SD)		Cumulative entropy	Mean level changes per choice task (SD)	
					Tasks	Alt.	Att.	Level	Level range					
16	83	1494	16.13	6.02	18	5	6	3	-20%	0.70	(0.39)	12.62	14.56	(1.57)
10	150	900	24.33	14.00	6	5	7	3	+20%	0.68	(0.38)	4.08	17.83	(1.46)
12	79	948	25.42	13.92	12	5	5	2	-20%	0.65	(0.39)	7.83	10.83	(1.07)
7	217	1302	28.73	15.67	6	4	7	2	-20%	0.51	(0.35)	3.05	13.04	(1.51)
1	83	1992	31.38	12.05	24	4	5	3	Base	0.56	(0.44)	13.43	9.33	(0.78)
8	79	948	33.33	11.39	12	5	4	4	+20%	0.91	(0.37)	10.91	11.50	(1.66)
9	80	1920	34.90	16.25	24	5	4	4	Base	0.98	(0.35)	23.47	10.88	(1.39)
13	111	1998	35.59	20.72	18	4	7	2	Base	0.40	(0.33)	7.21	9.89	(1.05)
2	81	1458	39.99	14.81	18	4	5	4	+20%	0.83	(0.35)	14.92	12.44	(2.27)
11	88	528	40.34	19.32	6	4	6	4	-20%	0.77	(0.28)	4.62	13.00	(1.41)
4	80	960	44.12	20.00	12	3	6	4	Base	0.45	(0.28)	5.37	9.33	(1.37)
14	81	1458	46.84	16.05	18	3	4	3	-20%	0.73	(0.33)	13.05	6.39	(0.95)
15	144	1728	50.46	25.69	12	3	5	2	Base	0.39	(0.27)	4.65	6.58	(0.95)
5	102	612	51.47	18.63	6	3	4	3	+20%	0.69	(0.15)	4.16	6.00	(0.58)
6	80	1920	53.80	15.00	24	3	4	4	-20%	0.79	(0.29)	18.94	7.29	(1.02)
3	123	2952	54.54	23.58	24	3	6	2	+20%	0.56	(0.21)	13.44	10.33	(1.18)
Total	1661	23118	Ø 37.84	Ø 16.92						Ø 0.66		Ø 10.12	Ø 10.58	

Note: Designs are ordered according to the share of status quo choices; N = observations; SQ = status quo; alt. = alternatives; att. = attributes; SD = standard deviation

The correlation patterns among the design dimensions and the structural complexity measures are reported in Table 6. Results show that cumulative entropy is strongly influenced by the number of choice tasks, and that the number of level changes is first of all influenced by the number of alternatives and then by the number of attributes per choice task. Compared to this the value of the entropy measure is only weakly associated with a single design dimension.

Table 6: Correlation patterns between design dimensions and structural complexity measures

Design dimension	Entropy	Cum. entropy	Number of level changes
Number of choice tasks	0.05 (0.40)	0.91 (0.00)	-0.11 (0.10)
Number of alternatives	0.20 (0.00)	0.10 (0.14)	0.64 (0.00)
Number of attributes	-0.33 (0.00)	-0.34 (0.00)	0.46 (0.00)
Number of levels	0.34 (0.00)	0.29 (0.00)	0.05 (0.46)
Level range	0.00 (0.99)	-0.06 (0.32)	0.17 (0.01)

Note: P-value in parentheses

4.4.3 Modelling results

Table 7 presents the estimates for the four models outlined in Table 3. In the first part of Table 7 estimates for the means and standard deviations corresponding to the attributes as well as the error components are reported. The second part of the table then reports the estimates for the interactions between the ASCsq and the variables capturing context effects. All dummy variables are effect coded (Bech and Gyrd-Hansen 2005). At the end of the table fit measures are reported.

We start with a brief description of the upper part of Table 7. For the plain MNL model we find that all coefficients are highly significant and have the expected sign. On average people want higher shares of forests in their surroundings, higher levels of biodiversity, and less land conversion. Also the coefficient for cost is negative and significant indicating that respondents were less likely to choose an alternative with a higher price. Regarding the MXL models (Model 2 to 4) we find this pattern to be present for all attributes across all models. The attribute coefficients are again all highly significant with the expected sign. Moving to the standard deviations we find that all except the one for Bio_other3 are highly significant as well. This indicates that tastes vary considerably across respondents, a finding that was expected. The literature shows that taste heterogeneity is in general present among respondents to stated preference surveys. Additionally, the interviews were conducted across Germany and therefore people live in different landscapes with, for instance, differing shares of forest cover or differing rates of land conversion. The significant standard deviations of the error components (EC2 to EC4) confirm that the hypothetical alternatives are perceived differently from the current situation across all designs. The values of the standard deviations as well as the t-values are very similar across models indicating that the correlation among the hypothetical alternatives is not influenced by the inclusion of other model components.

Turning to the lower part of Table 7, which reports the influence of the context variables on the probability to choose the status quo alternative, we see that the ASCsq in the MNL (Model 1) is highly significant with a positive sign. This indicates that participants are on average inclined to stay at the current situation. In this model also all interactions between the ASCsq and the socio-demographic variables are statistically significant. Thus, an advancing age as well as being female raises the probability to choose the status quo alternative while, in contrast, higher education increases the probability to choose one of the hypothetical alternatives. The ASCsq becomes indistinguishable from zero when we look at the remaining Models. This result is due to the more flexible assumptions of the MXL models. Model 2, 3 and 4 have not only random parameters, but also include the random error components which capture the heterogeneity among alternatives that is not attributable to taste heterogeneity with respect to the choice attributes.

Model 3 additionally incorporates a number of interaction terms with the ASCsq representing three types of context variables. These are the design dimensions, the structural complexity measure entropy, and the perceived status quo regarding the share of forest, land conversion, and biodiversity. Some of the interactions are highly significant indicating an important influence on the propensity to choose the status quo alternative. Starting with the design dimensions, we find that participants are more likely to choose the status quo alternative the later a choice task is presented in the sequence of tasks (position of a task). This result is in line with findings by Boxall et al. (2009), for example, providing further evidence that with an increasing number of decisions respondents are more likely to adopt simplifying strategies.

For the number of alternatives we find evidence, again in line with previous literature, for preference matching effects (e.g., Rolfe and Bennett 2009; Zhang and Adamowicz 2011). These effects are revealed through negative and significant coefficients the interactions indicating designs with four and five alternatives compared to designs with three alternatives. The more alternatives a choice task comprises, the less likely respondents are to choose the status quo. This association was already evident from Table 5 showing the relation between the share of status quo choices and the number of alternatives suggesting that with more alternatives it might have been easier for respondents to find a program that better matches their preferences. Thus, they are therefore less likely to select the status quo alternative.

Interestingly, status quo choices are unaffected by the number of attributes. At a first glance this result appears to be somewhat surprising since one might expect status quo choices to increase with more attributes as more information has to be processed by participants. An explanation for this result could be the specific design approach of this study. Here, the increase in the number of attributes in achieved by disaggregating the biodiversity in the whole landscape into its different components based on different landscape types. For instance, “biodiversity in the whole landscape” in designs with four attributes is split-up into “biodiversity in agricultural landscapes” and “biodiversity in other landscape types” in designs with five attributes. Thus, the number of attributes is varied without changing the good to be valued.

Also, no significant influence of the number of attribute levels on status quo choices is found. In contrast, the signs of the level range dummies, i.e., to have narrower or wider level ranges compared to the base level, indicate that participants are less inclined to choose the current situation when the level range is less extreme. A reason for this are probably non-compensatory preferences or attribute cut-offs (Swait 2001). When the level range becomes wider, some level values may exceed the maximum (minimum) acceptable level. As a consequence, respondents remained at the status quo to avoid cutoff violations.

Regarding the structural complexity measures, the interaction term of the ASCsq and entropy is found to be highly significant suggesting that respondents opted more often for the status quo when the similarity between alternatives was high and therefore the decision more difficult. This result is similar to other findings presented in the discrete choice literature (Balcombe and Fraser 2010; Zhang and Adamowicz 2011; Campbell et al. 2015a). As already stated, the other two structural measures, cumulative entropy and the number of level changes, have not been incorporated due to high correlations with the corresponding design dimensions choice task number and number of alternatives. In both cases the correlation exceeds 0.5 (Table 6). Finally, accounting for the influence of the perceived current situation on the probability to choose the status quo alternative, we find a significant effect on status quo choices. The higher the perceived share of forests and biodiversity in the landscape surrounding respondents’ place of residence, the higher the probability to opt for the status quo alternative. The opposite effect is observed for the interaction with land conversion. We interpret this finding as an indication for the role preferences play for the decision to opt for the status quo alternative or not. The three interactions reveal that the higher respondents perceived the quality of the landscape in their surroundings, the more likely they are to choose the status quo.

Table 7: Model estimates (Coefficients and robust |t-value|)

	Model 1 MNL			Model 2 MXL			Model 3 MXL			Model 4 Scale MXL		
	Coefficient	|rob. t-value|		Coefficient	| rob. t-value |		Coefficient	| rob. t-value |		Coefficient	| rob. t-value|	
Attributes												
Share of Forest	0.0169	35.67	***	0.0314	22.23	***	0.0309	20.23	***	0.0383	4.29	***
Land Conversion	-0.0086	38.20	***	-0.0160	19.01	***	-0.0156	19.61	***	-0.0192	4.22	***
Bio_whole	0.0097	8.69	***	0.0149	4.52	***	0.0159	4.76	***	0.0192	3.17	***
Bio_agra	0.0056	9.47	***	0.0086	6.82	***	0.0079	5.77	***	0.0110	3.58	***
Bio_forest	0.0058	6.49	***	0.0143	7.41	***	0.0142	7.37	***	0.0201	3.59	***
Bio_urban	0.0058	5.32	***	0.0108	5.75	***	0.0098	5.17	***	0.0167	3.35	***
Bio_other1	0.0077	7.01	***	0.0112	3.85	***	0.0097	3.47	***	0.0135	2.89	***
Bio_other2	0.0029	4.02	***	0.0096	6.19	***	0.0103	6.17	***	0.0127	3.49	***
Bio_other3	0.0050	5.45	***	0.0084	5.55	***	0.0078	5.14	***	0.0125	3.29	***
Cost	-0.0062	25.52	***	-5.3500	32.15	***	-5.5500	31.87	***	-5.2800	20.35	***
Standard deviations												
Share of Forest				0.0357	23.22	***	0.0352	20.87	***	0.0408	4.36	***
Land Conversion				0.0220	25.40	***	0.0220	24.87	***	0.0257	4.30	***
Bio_whole				0.0370	6.43	***	0.0353	9.39	***	0.0472	3.71	***
Bio_agra				0.0125	5.10	***	0.0114	3.94	***	0.0170	3.72	***
Bio_forest				0.0159	6.17	***	0.0166	6.00	***	0.0202	3.68	***
Bio_urban				0.0129	3.98	***	0.0130	3.78	***	0.0209	3.02	***
Bio_other1				0.0261	6.06	***	0.0268	6.12	***	0.0327	3.60	***
Bio_other2				0.0120	6.38	***	0.0141	5.64	***	0.0143	3.21	***
Bio_other3				0.0046	1.19		0.0052	1.03		0.0087	1.08	
Cost				2.8200	24.97	***	2.4600	18.23	***	3.0100	21.52	***
Error components												
EC_2				3.8900	13.74	***	3.8400	15.78	***	4.5700	4.42	***
EC_3				4.5500	13.46	***	4.9500	15.20	***	4.9900	3.70	***
EC_4				4.1200	13.00	***	4.8600	10.97	***	5.3100	4.04	***
Alternative specific constant												

ASCsq	1.0200	13.90	***	0.0642	0.08		2.5900	1.61		-2.0500	-1.40	
Socio-demographic variables												
× age	0.0029	2.79	***	0.0117	0.88		0.0122	0.70		0.0061	0.52	
× gender	-0.2990	10.71	***	-0.3090	2.48	**	-0.4950	1.70	*	-0.2650	-1.72	***
× education	-0.1630	11.31	***	-0.4410	3.51	***	-0.2510	-1.82	*	-0.5590	-2.16	***
Structural complexity measure												
× entropy							0.3650	4.08	***	0.3020	2.40	**
Perceived status quo												
× share of forest							0.7130	2.85	***	1.0600	2.45	***
× land conversion							-0.9760	-2.42	**	-0.8810	-2.69	***
× biodiversity							0.7510	2.67	***	0.6120	2.49	**
Design dimensions in												
… utility function												
× position of task							0.0472	7.30	***			
× number of alternatives							-1.0900	-6.98	***			
× number of attributes							-0.1300	-0.95				
× number of levels							-0.0030	-0.01				
× narrow level							-0.3330	-2.72	***			
× wide level							0.5680	1.66	*			
… scale function												
-> position of task										0.0214	7.74	***
-> number of alternatives										-0.0381	-1.18	
-> number of attributes										-0.0380	-1.39	
-> number of levels										-0.0134	-0.37	
-> narrow level										0.0971	3.14	***
-> wide level										-0.1760	-5.66	***
Log-likelihood	-26870.47			-17682.68			-17560.32			-17520.66		
Number of parameters	14			27			37			37		
Observations	23118			23118			23118			23118		
Akaike Information Criterion	53769			35419			35195			35115		
Bayesian Information Criterion	53882			35637			35492			35413		

Note: ***, **, *: significance at the 1%, 5% and 10% level, × interaction with the ASCsq, -> covariate in scale function

The final model (Model 4) incorporates, in contrast to Model 3, the design dimensions into the scale function, but keeps everything else unchanged. Regarding the context variables that remain in the utility function, we find that the effects are mainly the same as in Model 3. Entropy as well as the interaction terms of the ASCsq with the perceived status quo are still significant and have the same sign.

When we compare the effects of the design dimensions in the utility function to those in the scale function, we find that three variables, position of task, narrow as well as wide level range, are statistically significant in both models. The variance of the error term decreases with an increasing number of choice tasks indicating some institutional learning reported also in other studies (e.g., Czajkowski et al. 2014). Choice consistency is furthermore promoted when level ranges are narrow, but decreases when the level range broadens. When the range is narrow comparisons of alternatives are easier for respondents. This finding is in line with Caussade et al. (2005). The other design dimensions, including the number of alternatives, which has been found to influence status quo choices significantly in Model 3, do not significantly impact choice consistency.

Regarding model performance we find, not surprisingly, that all MXL models clearly outperform the MNL model. A comparison of the log-likelihood values and both information criteria for Model 1 and Model 2 suggests that accounting for unobserved taste heterogeneity with respect to attributes and alternatives as well as for the panel character of the data strongly improves model fit. Incorporating the context variables again improves the fit when Model 3 is compared to Model 2. The best model fit is achieved in Model 4 where the design dimensions are covariates of the scale parameter; although Model 4 has the same number of parameters, the log-likelihood is lower than in Model 3. This has likely to do with the distributional assumption underlying the random parameters of these models. When the design dimensions are included in the scale function, complexity impacts are modeled in a more flexible way since the distribution can change among individuals, that is, the distributional assumptions in Model 4 are more flexible than in Model 3.

4.4.4 Effects on welfare

Subsequently, we investigate whether different designs result in different marginal WTP estimates and compensating variation (CV) measures. According to Table 5 the designs are associated with different shares of status quo choices. Since we expect that potential differences in the non-marginal CV measures are driven by the sign and value of the ASCsq rather than by the marginal WTP estimates, we select the two designs with the lowest (16.13%) and the highest

(54.54%) share of status quo choices, i.e., Design 16 and Design 3, respectively. By chance both designs have the same number of attributes and differences in welfare measures should therefore not be caused by a varying number of attributes used to describe the changes in biodiversity. For both designs we estimate a MNL and a MXL model following the specification of Model 2 in Table 3 (model results are presented in Appendix A). Both models are estimated, as stated in the econometric section, in WTP space to avoid extremely high WTP values and to obtain more reasonable WTP distribution. In order to calculate the CV measures we assume a scenario in which the share of forests increases by 10% and the rate of land conversion decreases by 5%. For biodiversity we assume, for the sake of simplicity, no changes compared to the status quo. As Adamowicz et al. (2011) point out, the literature provides little guidance on whether the status quo parameter should be included or excluded from non-marginal welfare calculations. Therefore, Table 8 presents the welfare measures from both models each time with and without accounting for the status quo effect via including the ASCsq or not.

Table 8: Marginal willingness to pay and compensating variation measures based on WTP space models in Euro

Model			Design 3 (24 Tasks, 3 Alt., 6 Att.)	Design 16 (18 Tasks, 5 Alt., 6 Att.)
MNL				
Marginal WTP		Share of Forest	1.30 (0.66 / 1.94)	3.52 (2.36 / 4.68)
		Land Conversion	-0.80 (-1.04 / -0.56)	-1.96 (-2.44 /-1.48)
Compensating variation	ASCsq	Yes	106.01 (44.83 / 167.19)	-172.00 (-220.00 / -123.99)
		No	-16.99 (-24.24 / -9.74)	-45.07 (-58.28/ -31.72)
MXL				
Marginal WTP		Share of Forest	1.09 (0.62 / 1.56)	5.53 (2.20 / 8.86)
		Land Conversion	-0.87 (-3.13 / 1.41)	-2.98 (-4.93 / -1.02)
Compensating variation	ASCsq	Yes	129.78 (55.75 / 203.79)	-705.20 (-1148.77 / -261.63)
		No	-15.23 (-20.47 / -9.98)	-70.20 (-106.44 / -33.95)

Note: Confidence interval (in parentheses) calculated using the Delta method; WTP = willingness to pay (€ per household and year) MNL = Multinomial Logit; MXL = Mixed Logit

As reported in Table 8, estimates obtained from the MNL as well as the MXL lead to higher mean marginal WTP estimates for Design 16 compared to Design 3. Respondents who have been randomly assigned to Design 16 are willing to pay more for a one percent increase in forest share (€3.52 compared to €1.30 for the MNL and €5.53 compared to €1.09 for the MXL) and experience a larger disutility from a one percent increase in land conversion (€-1.96 compared to €-0.80 for the MNL and €-2.98 compared to €-0.87 for the MXL). Considering the CV for the land use change scenarios presented above, the pattern remains the same. The non-marginal values for Design 16 indicate much larger welfare gains than those derived from Design 3. As stated above, we distinguish here between measures calculated including the value of the ASCsq and excluding it. Starting with the latter, we find values which are an order of magnitude smaller. Based on Design 16 and the MNL model, €-45 are needed to keep individuals at the same utility level compared to €-17 based on Design 3. This is a difference of around €30. The confidence intervals are non-overlapping. The difference between the designs becomes even larger when we consider the results based on the MXL model. Here, the values are calculated to be €-70 for Design 16 and of €-15 for design 3; the difference has increased to €55.

Proceeding to the CV measures that include the value of the ASCsq, the differences increase strongly to around €825. Confidence intervals are again both times non-overlapping indicating statistically significant differences. It is also noteworthy that when we include the ASCsq in calculating the non-marginal values the sign flips from positive in Design 3 to negative in Design 16. If we use Design 3, results suggest that individuals would on average experience a loss when the current situation would be changed while Design 16 suggests the opposite. Finally, we tested all comparisons using the complete combinatorial method (Poe et al. 2005). In all cases we can clearly reject the null hypothesis that the mean values are from the same distribution.

4.5 Conclusions

To the best of our knowledge this is the first attempt to examine the status quo effect in a DoD approach systematically accounting for five design dimensions of the CE. Particularly, in environmental and resource economics knowing whether status quo choices are triggered by the design of the experiment is crucial. In contrast to other fields such as transportation or health economics, the option to stay with the status quo is often a reasonable response as people might prefer the status quo and do not want to give up income for improvements or to avoid deteriorations. If the propensity to choose the status quo alternative, however, is strongly influenced by the design of the CE, subsequent welfare estimates might be biased. Besides the design dimensions of the CE, we further incorporate context variables such as structural complexity measures and the perceived status quo to control for their influence on model results.

One of our main findings is that the frequency of status quo choices is negatively affected by the number of alternatives likely due to preference matching. Respondents who are presented only a few alternatives are more likely to select the status quo alternative as it might be more difficult for them to find an alternative that matches their preferences (Zhang and Adamowicz, 2011). This result is worrying since we find the highest number of status quo choices in designs with two hypothetical and one status quo alternative. According to our understanding of the literature, this choice task format is the most often used in environmental economics. If our results are generalizable, this suggests that high frequencies of status quo choices reported in other studies might have been induced by the design of the choice task. In contrast, we do not find any evidence for a positive association between the number of alternatives and the error variance.

We further find support for the hypothesis that status quo choices are positively related to the number of choice tasks respondents face. Presenting, everything else equal, more choice tasks is likely to increase the number of status quo choices. When we investigate the impact of the design dimensions on scale, we observe the error variance to decrease with an increasing number of choice tasks. This is in line with institutional learning (Czajkowski et al., 2014, Carlsson et al., 2012, Day et al., 2012, Hess et al., 2012), but might be the result of adopting non-compensatory decision heuristics, too. This issue is also raised by Czajkowski et al. (2014) as an alternative explanation for decreases in the error variance which is also found in their study. Therefore, the decreasing error variance we observe might not only be explained by institutional learning, but could also be due to the fact that some respondents adopted non-compensatory decision heuristics, e.g. to choose the status quo option more frequently towards the end of the CE. Indeed, Campbell et al. (2015b) show that fatigue and learning effects vary across respondents. Moreover, note that because of identification problems we are not able to investigate the impact of the design dimensions on both utility and scale simultaneously (see also Zhang and Adamowicz, 2011).

Regarding the number of attributes, however, it has to be kept in mind that the observed effects may be due to the specific design approach we use in this study. Increases in the number of attributes have been achieved by disaggregating one attribute, i.e., the attribute concerned with biodiversity. We employed this approach in order to be able to systematically account for differences in WTP estimates with respect to a common-metric attribute (Hensher, 2004, Meyerhoff et al., 2015). As a consequence, respondents may have jointly evaluated the biodiversity attributes rather than considering each in isolation. Alternatively, participants might not have attended to all components of the common-metric biodiversity attribute (Hensher and

Greene 2010). Both approaches could be seen as a simplifying information processing strategy and might therefore be responsible for our findings.

For the remaining design dimensions we find the level range to be positively associated with the number of status quo choices as well as the error variance. We assume that the reason for this result is that larger level ranges signal to respondents a higher degree of uncertainty with respect to the magnitude of the environmental change to be valued.

Another effect on status quo choices results from the levels of similarity between alternatives. Similarity, measured through the degree of entropy, is found to positively influence status quo choices. This shows that the findings by Swait and Adamowicz (2001) and Zhang and Adamowicz (2011), who found that higher entropy makes a status quo option more attractive, apply to a wider range of choice task formats. The other structural complexity measures, cumulative entropy and the number of attribute level changes, are highly correlated with the design dimensions and it is therefore not possible to independently determine their influence on the propensity to choose the status quo alternative. Status quo choices are also influenced by preferences. This is revealed through a positive relationship between the perceived quality of the current situation in respondents' surroundings and the likelihood to stay at the status quo. This finding indicates the validity of the results as it is in line with demand theory. People who perceive the environmental quality in their surroundings as higher are expected to experience diminishing utility from further quality improvements.

Effects triggered by the design of the choice tasks would be of less concern when they had no impact on welfare measures. Unfortunately, our results strongly suggest the opposite. Differences in the design of the CE are observed to have significant influence on marginal as well as non-marginal welfare estimates. Calculations for the two designs that cause the lowest and the highest number of status quo choices result for both, marginal and non-marginal estimates, in statistically different estimates. By including or excluding the status quo parameter in the calculations of the non-marginal welfare measures, we even observe a change in the sign, i.e., individuals experience utility from the valued land use changes according to one design, but would experience a loss of utility from the same land use changes according to the other design. Similar findings were presented, among others, by Adamowicz et al. (1998) and Boxall et al. (2009). This result indicates the need for further investigations and emphasizes at the same time the need for guidelines on whether the ASCsq should be considered or not when calculating non-marginal welfare measures.

Based on our results we would like to give at least two recommendations to researchers and practitioners who have to design a CE. First, we suggest considering to employ a design that goes beyond choice task formats with two hypothetical and a status quo alternative. Increasing the number of alternatives offers respondents more options resulting in a higher probability of preference matching. Second, we suggest that the number of choice tasks should be of minor concern when generating a CE design. In many studies authors report that the number of choice tasks was limited to, for example, four or six tasks per respondent in order to avoid negative effects of fatigue. The recent literature (Czajkowski et al., 2014, Hess et al., 2012, Meyerhoff et al., 2015) as well as the results of this study both do not support such a conclusion. Thus, a number of choice tasks in the range between 10 and 15 should in most applications not affect study results negatively. A larger number of choice tasks is at the same time desirable for calculating individual specific WTP estimates (e.g., Train, 2009). Another interesting result is that designs with a higher number of attributes do not seem to impair responses. This also supports the use of larger choice tasks. However, as detailed above, the DoD approach might have promoted this finding and it thus has to be interpreted with a higher degree of caution.

Future research applying a DoD approach should consider the inclusion of choice tasks with only one hypothetical and a status quo alternative. As this design is thought to be incentive compatible (Carson and Groves, 2007, Collins and Vossler, 2009), including such tasks would enable further interesting insights. If possible, future DoD studies might also aim at increasing sample sizes. This could be especially crucial when it comes to calculating welfare measures based on separate designs. Our findings are restricted to the comparison of two sub-samples with 123 and 83 respondents, respectively. Finally, one might want to vary the number of attributes by a different approach than the one followed here. The problem, however, is not to confound effects of an increased number of attributes with effects due to information conveyed by attributes.

Acknowledgments

The authors would like to thank three anonymous reviewers for their comments and suggestions. Moreover, we would like to thank Vic Adamowicz and the participants of the seminar at the Department of Resource Economics & Environmental Sociology at the University of Alberta, Edmonton, Canada, for very fruitful discussions and valuable suggestions during a stay of Malte Oehlmann and Jürgen Meyerhoff at the department in autumn 2014. The research stay of Malte Oehlmann was supported by the German Academic Exchange Service "DAAD Doktoranden Stipendium". Petr Mariel would like to acknowledge financial support from Basque Government through grant IT-642-13, Spanish Ministry of Economy and Competitiveness through grant ECO2014-52587-R. This paper was finalized when Petr Mariel was visiting the Durham University Business School. He gratefully acknowledges the support provided by the Basque Government (Ikermugikorta-suna 2016) for this stay. Funding for this research was provided by the Federal German Ministry of Education and Research (Fkz. 033L029G; Fkz01LL0909A) and is gratefully acknowledged as well.

4.6 References

Adamowicz, W.L., Boxall, P., Williams, M., 1998. Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. American Journal of Agricultural Economics, 80(1), 64–75.

Adamowicz, W., Dupont, D., Krupnick, A., Zhang, J., 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. Journal of Environmental Economics and Management, 61(2), 213–226.

Balcombe, K.G., Fraser, I., 2010. A General Treatment of 'Don't Know' Responses from Choice Experiments. European Review of Agricultural Economics, 38(2), 171–191.

Bech, M., Gyrd-Hansen, D., 2005. Effects coding in discrete choice experiments. Health Economics, 14, 1079–1083.

Bennett, J., Adamowicz, V., 2001. Some Fundamentals of Environmental Choice Modelling. In: Bennett, J., Blamey, R., eds.: The Choice Modelling Approach to Environmental Valuation. Cheltenham, 37-71.

Beshears, J., Choi, J.J., Laibson, D., Madrian, B.C., 2008. How are preferences revealed? Journal of Public Economics, 92(8-9), 1787–1794.

Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models, in: Chevroulet, T., Sevestre, A. (Eds.), Proc. 3rd Swiss Transportation Research Conf., 19-21 March 2003, Monte-Verita, Ascona, Switzerland.

Bierlaire, M., 2008. An Introduction to BIOGEME Version 1.7. Available at: biogeme.epfl.ch.

Bliemer M. C. J., Rose J. M., 2011. Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transportation Research Part A: Policy 45, 63-79.

Boxall, P., Moon, A., Adamowicz, W.L., 2009. Complexity in Decision Making: The Effect of Task Structure and Context on Participant Response. Australian Journal of Agricultural and Resource Economics, 53(4), 503–519.

Bullock, C. H. 2008. Valuing urban green space: Hypothetical alternatives and the status quo. Journal of Environmental Planning and Management, 51(1), 15-35.

Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), 2010. Der Indikatorenbericht 2010 zur Nationalen Strategie zur biologischen Vielfalt. Berlin.

Campbell, D., Mørkbak, M.R., Olsen, S.B., 2015a. Assessing the impact of opt-out definitions in hoice experiments - A case study on food safety. The forth International Choice Modelling Conference 2015, Austin, Texas, 10-13 May 2015.

Campbell, D., Boeri, M., Doherty, E., Hutchinson, G., 2015b. Learning, fatigue and preference formation in discrete choice experiments. Journal of Economic Behavior & Organization, 119, 345-363.

Carlsson, F., Mørkbak, M.R., Olsen, S.B., 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling, 5, 19-37.

Carson, R.T., Groves, T., 2007. Incentive and Informational Properties of Preference Questions. Environmental and Resource Economics, 37, 181–210.

Caussade, S., Ortúzar, J., Rizzi, L. I., Hensher, D. A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B: Methodological, 39, 621-640.

Cherchi, E., Hensher, D. A., 2015. Workshop synthesis: Stated preference surveys and experimental design, an audit of the journey so far and future research perspectives. Transportation Research Procedia, 11, 154-164.

ChoiceMetrics, 2012. Ngene 1.1.1 User manual and reference guide. http://www.choicemetrics.com.

Collins, J.P., Vossler, C.A., 2009. Incentive compatibility tests of choice experiment value elicitation questions. Journal of Environmental Economics and Management, 58(2), 226–235.

Czajkowssk, N., Giergiczny, M., Greene, W.H., 2014. Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. Land Economics, 90, 324-351.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P., 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management, 63, 73-91.

DeShazo, J.R., Fermo, G., 2002. Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management, 44, 123–143.

Dhar, R., 1997. Consumer Preference for a No-Choice Option. Journal of Consumer Research, 24(2), 215–231.

Fosgerau, M., Börjesson, M., 2015. Manipulating a stated choice experiment. Journal of Choice Modelling, 16, 43-49.

Hanley, N., Wright, R. E., Alvarez-Farizo, B., 2006. Estimating the economic value of improvements in river ecology using choice experiments: An application to the water framework directive. Journal of Environmental Management, 78(2), 183-193.

Hartman, R.S., Doane, M.J., Woo, C.-K., 1991. Consumer Rationality and the Status Quo. The Quarterly Journal of Economics, 106(1), 141–162.

Hensher, D.A., 2004. Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. Journal of Transport Economics and Policy 38: 425-446.

Hensher, D.A., 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics, 21, 861–878.

Hess, S., Beck, M. Chorus, C., 2014. Contrasts between utility maximisation and regret minimisation in the presence of opt out alternatives. Transportation Research Part A: Policy and Practice, 66, 1–12.

Hess, S., Hensher, D.A., Daly, A., 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: Policy and Practice, 46(3), 626-44.

Hess, S., Rose, J.M., 2009. Should Reference Alternatives in Pivot Design SC Surveys be Treated Differently? Journal of Environmental and Resource Economics, 42(3), 297–317.

Hole, A. R., 2007. Fitting mixed logit models by using maximum simulated likelihood. Stata Journal 7(3), 388-401.

Kahneman, D., Knetch, J.L., Thaler, R.H., 1991. Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. The Journal of Economic Perspectives, 5(1), 193–206.

Kataria, M., Bateman, I., Christensen, T., Hasler, B., Hime, S., Ladenburg, J., Levin, G., Martinsen, L., Nissen, C., 2012. Scenario realism and welfare estimates in Choice Experiments – A Non-market Valuation Study on the European Water Framework Directive. Journal of Environmental Management, 94(1), 25–33.

Lanz, B., Provins, A. 2015. Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? Journal of Regulatory Economics, 47(3), 300-324.

Marsh, D., Mkware, L., Scarpa, R., 2011. Do Respondents' Perceptions of the Status Quo Matter in Non-Market Valuation with Choice Experiments? An Application to New Zealand Freshwater Streams. Sustainability, 3(9), 1593–1615.

Meyerhoff, J., Oehlmann, M., Weller, P., 2015. The Influence of Design Dimensions on Stated Choices in an Environmental Context. Environmental and Resource Economics, 61(3), 385-407.

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational Methods for Measuring the Difference of Empirical Distributions. American Journal of Agricultural Economics 87, 353-365.

Ritov, I., Baron, J., 1994. Outcome Knowledge, Regret, and Omission Bias. Organizational Behavior and Human Decision Processes, 64(2), 119–127.

Rolfe, J., Bennett, J., 2009. The impact of offering two versus three alternatives in choice modelling experiments. Ecological Economics, 68(4), 1140–1148.

Rolfe, J., Windle, J., 2012. Restricted versus Unrestricted Choice in Labeled Choice Experiments: Exploring the Trade-Offs of Expanding Choice Dimensions. Canadian Journal of Agricultural Economics, 60(1), 53–70.

Samuelsson, W., Zeckhauser, R., 1988. Status Quo Bias in Decision Making. Journal of Risk and Uncertainty, 1, 7–59.

Scarpa, R., Ferrini, S., Willis, K.G., 2005. Performance of Error Component Models for Status-Quo Affects in Choice Experiments. In Applications of Simulation Methods in Environmental and Resource Economics, ed. Scarpa, R. and Alberini, A. Dordrecht, Netherlands: Springer.

Scarpa, R., Willis, K.G., Acutt, M., 2007. Valuing externalities from water supply: Status quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. Journal of Environmental Planning and Management, 50(4), 449–466.

Scarpa, R., Rose, J., 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. The Australian Journal of Agricultural and Resource Economics, 52, 253–282.

Shafir, E., 1993. Choosing versus rejecting: Why some options are both better and worse than others. Memory & Cognition, 21(4), 546–556.

Shannon, C.E., 1948. A note on the concept of entropy. Bell System Technical Journal, 28, 379.

Swait, J., Adamowicz, W.L., 2001. Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. Organizational Behavior and Human Decision Processes, 86(2), 141–167.

Swait, J., 2001. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodigical, 35(10), 903‑928.

Train, K., 2009. Discrete choice methods with simulation. Cambridge University Press.

Tversky, A., Shafir, E., 1992. Choice Under Conflict: The Dynamics of Deferred Decision. Psychological Science, 3(6), 358–361.

Von Haefen, R.H., Massey, D.M., Adamowicz, W.L., 2005. Serial Nonparticipation in Repeated Discrete Choice Models. American Journal of Agricultural Economics, 57(4), 1061–1076.

Willis, K.G., 2009. Assessing visitor preferences in the management of archaeological and heritage attractions: a case study of Hadrian's Roman Wall. Journal of Tourism Research, 11(5), 487–505.

Yao, R.T., Scarpa, R., Rose, J.M. and Turner, J.T. (2015). Experimental Design Criteria and Their Behavioural Efficiency: An Evaluation in the Field. Environmental and Resource Economics, 62, 433-455.

Zhang, J., Adamowicz, W.L., 2011. Unraveling the Choice Format Effect: A Context-Dependent Random Utility Model, Land Economics, 87(4), 730 – 743.

4.7 Appendix

Appendix A: WTP space models for Design 3 and Design 16

Variable	Design 3 (24 Tasks, 3 Alt., 6 Att.)				Design 16 (18 Taskss, 5 Alt., 6 Att.)			
	MNL		MXL		MNL		MXL	
	Coefficient	|rob. t.|	Coefficient	|rob. t.|	Coefficient	|rob. t.|	Coefficient	|rob. t.|
Means								
ASCsq	123.000	3.56 ***	145.000	3.75***	-127.000	5.07***	-635.000	2.85***
Share of Forest	1.300	3.93***	1.090	4.58***	3.520	5.93***	5.530	3.25***
Land conversion	-0.798	6.47***	-0.865	4.78***	-1.960	8.01***	-2.980	2.98***
Bio_agrar	0.498	2.40***	0.323	2.68***	-1.350	1.43***	-1.240	2.02**
Bio_forest	1.000	3.15***	0.668	3.96***	-2.680	2.60***	-2.600	2.46**
Bio_other2	0.821	3.53***	0.606	3.62***	0.157	0.30***	0.346	0.36
Cost	-0.006	6.27***	-3.920	25.24***	-0.007	8.78***	-4.780	10.21***
Standard deviations								
Share of Forest			1.250	5.05***			6.560	2.33***
Land Conversion			1.160	4.94***			3.290	3.49***
Bio_agrar			0.219	0.88***			0.003	0.01
Bio_forest			0.689	3.04***			1.910	0.33
Bio_other2			0.554	3.72***			1.740	0.96
Cost			0.707	7.62***			0.879	2.96***
Error components								
EC2			334.000	4.31***				
EC3							680.000	2.43**
Number of parameters	7		14		7		14	
Observations	2952		2952		1494		1494	
Log-likelihood	-2739.97		-1505.09		-2117.65		-1638.22	
AIC	5493.94		3036.19		4249.29		3304.43	
BIC	5535.87		3120.05		4286.46		3378.76	

Note: rob. t. = robust t-values; AIC =Akaike Information Criterion; BIC =Bayesian Information Criterion

5 Do Attribute Cut-offs Make a Difference? The Effects of Eliciting and Incorporating Cut-off Values in Choice Models

Malte Oehlmann[a]

Abstract

Despite theoretical assumptions, empirical evidence suggests that individuals may adopt non-compensatory decision strategies such as attribute thresholds (cut-offs) when having to make choices. This paper analyzes the effect of eliciting and incorporating attribute cut-offs in choice models. Using data from a large-scale survey on the development of renewable energies in Germany, I first investigate whether cut-off elicitation questions prior to a choice experiment influence preferences. Using a split sample approach, I show that, compared to the treatment where no attribute thresholds are reported, willingness to pay estimates are higher when cut-offs are elicited. In the next step, the subsample with cut-off elicitation is further analyzed. Descriptive statistics show that most participants state a cut-off level for attributes. However, respondents are willing to violate their self-imposed thresholds when making trade-offs in the subsequent choice experiment. To account for this effect, stated cut-offs are incorporated into a latent class model following the soft cut-off approach proposed by Swait (2001). I find that including cut-offs in the choice model improves model fit and effects the estimated parameters.

Keywords: choice experiment, decision heuristics, attribute cut-offs, willingness to pay

Working paper presented at the 5th International Choice Modelling conference 2017 and the 6th World Congress of Environmental and Resource Economists

[a] Technische Universität Berlin, Institute for Landscape Architecture and Environmental Planning, Straße des 17. Juni 145, D-10623 Berlin, Germany

5.1 Introduction

Renewable energy development has been rapidly growing in recent years. Perhaps nowhere in the world is this more evident than in Germany where government policies such as the German Renewable Energy Act has increased the share of energy from renewable sources in electricity production to 32 percent in 2016 (BMWi 2017). This trend is expected to continue in the future. By 2025, the share of renewable energies in electricity production is aimed to obtain 40-45% (BMWi 2017). To incorporate new renewable energy sources into the electricity system the need for expanding the transmission grid has been highlighted in several instances (e.g. Schroeder et al. 2013). Despite public support of renewable energy technologies in Germany, the construction of new renewable energy facilities (REFs) as well as new transmission lines often faces substantial opposition at the local level. For instance, people living near wind farms may experience significant negative external effects from changes in the landscape, noise emission or shadow cast. Understanding the publics' preferences towards renewable energy alternatives and quantifying the social costs of renewable energy technologies is crucial to facilitate the successful and efficient expansion of renewable energies.

Discrete choice experiments (DCEs) have been found to be particularly suited to achieve these goals (Bergmann et al. 2008, Kosenius and Ollikainen, 2013). They have been applied in several studies to the economic valuation of preferences and willingness to pay (WTP) estimates towards different energy sources with a particular focus on specific onshore (e.g. Álvarez-Farizo and Hanley 2002, Dimitropoulos and Kontroleon 2009, Strazzera et al. 2011) and offshore wind sites (e.g., Ladenburg 2009, Krueger et al. 2011, Ladenburg et al. 2012). DCEs have also been employed to examine impacts of different sources simultaneously (e.g. Bergmann et al. 2008, Cicia et al. 2012, Kosenius and Ollikainen 2013) and to quantify negative externalities from transmission lines (McNair et al. 2011, Hu and Yoo 2014).

As is typically in choice modeling, these studies assume that individuals have compensatory decision strategies such that they make trade-offs among all attributes. However, empirical evidence suggests that individuals may adopt non-compensatory decision strategies to, for instance, manage information acquisition and processing cost (e.g. Swait 2001). Two well-documented heuristics are elimination-by-aspects (Tversky 1972) and satisficing (Simon 1955). A decision-making process that is associated with these heuristics is the use of thresholds or attribute cut-offs (Leong and Hensher 2012), which are defined as the minimum (maximum) acceptable level an individual defines for an attribute (Huber and Klein 1991). Swait (2001) and

Ding et al. (2012) both find that failing to account for such attribute cut-offs may lead to biased model parameter estimates and welfare measures.

The use of attribute cut-offs has been investigated in several areas of research including transportation (Danielis and Marcucci 2007, Marcucci and Gatta 2011, Hensher and Rose 2012, Roman et al. 2017), health economics (Mentzakis et al. 2011) and food choice (Ding et al. 2012, Moser and Raffaelli 2012, Moser and Raffaelli 2014, Peschel et al. 2016). To the best of my knowledge, there is only one attempt in witch attribute cut-offs were analyzed in an environmental context (Bush et al. 2009).

In principle, the use of attribute cut-offs may be investigated by following an analytical or a self-stated approach (Hensher 2013). Since the analytical method – a two-stage estimation approach (e.g. Manski 1977) - involves several technical challenges (e.g., Swait 2001, Ding et al. 2012), only the self-stated approach has been implemented. Here, respondents have been asked to state their attribute thresholds prior or after the choice tasks. Swait (2001) discusses merits and drawbacks of these approaches opting for eliciting attribute cut-offs prior to the DCE arguing that thresholds should be based on individuals' past experience and not on information provided in the choice tasks. I contribute to this literature by systematically addressing whether prior cut-off elicitation influences individuals' preferences in the subsequent choice tasks.

This paper investigates the use of attribute cut-offs in an environmental context using data from a labeled DCE on the development of renewable energies and long-distance transmission lines (LDTLs) in Germany. The DCE was part of a large-scale online survey conducted in 2013. Prior to the DCE, cut-offs were elicited from a split sample of 1,694 out of a total of 3,390 respondents. Respondents were randomly assigned to one of two treatments which only differed in whether cut-offs were stated or not. The first objective of this research is to systematically investigate whether cut-off elicitation questions prior to a DCE influence choices, i.e. preference parameters and subsequent WTP estimates. The second objective is to analyze whether participants of DCE stick to their self-stated constraints or whether they are willing to violate them in the choice process. Cut-off violations may occur because when considered in isolation, the minimum or maximum acceptable level of each attribute may actually reflect decision makers' purpose. However, when traded-off against other attributes, individuals may be willing to either change or violate cut-offs because the additional benefit is greater than the cost caused by the violation recognizing the opportunity cost of self-reported cut-offs (Swait 2001, Bush et al. 2009).

The violation of thresholds leads to the soft cut-off approach first proposed by Huber and Klein (1991) who found that individuals violate their stated cut-offs and adjust their thresholds when they have more information about the attributes and decision task. A model recognizing the idea of soft cut-offs expanding the random utility framework by introducing utility penalties when self-imposed thresholds are violated is Swait (2001). The Swait (2001) model is the approach that has been employed most frequently to account for cut-offs in a discrete choice framework (Bush et al. 2009, Ding et al. 2012, Peschel et al. 2016, Roman et al. 2017). However, research following the idea of soft cut-offs in the context of unobserved preference heterogeneity and heterogeneous response to attribute cut-offs is still scarce (see Ding et al. 2012, Peschel et al. 2016 and Roman et al. 2017 for some exceptions). I add to this research by specifying a Mixed Logit Model (MXL) with utility penalties for violated attributes. Parameter estimates then reveal whether individuals tend to adopt compensatory or non-compensatory decision rules (Swait 2001, Moser and Raffaelli 2012, Roman et al. 2017). The results of this research are expected to improve the understanding of decision-making with respect preferences towards renewable energies and the use of attribute cut-offs in an environmental valuation context.

5.2 Survey Design and Cut-off Elicitation

5.2.1 Study Design

The survey aimed at analyzing attitudes, preferences and acceptance towards the development of renewable energies in Germany. First, focus groups in six different cities throughout Germany were conducted in October 2012 to test the survey instrument. Among other things, participants were asked to answer and give feedback on parts of the questionnaire including the DCE. A revised questionnaire was then tested in two pilot studies. The final version of the questionnaire started by visualizing the renewable energy sources to be considered in the survey using pictograms.

Respondents were told that the survey focused solely on renewable energies in the open landscape and did not consider offshore wind power and solar panels installed on roofs. Hereafter, participants were asked to answer several warm-up questions such as their exposure to REFs, or whether they feel disturbed by REFs in their surroundings. At the end of the questionnaire several socio-demographic characteristics were requested. Six attributes were used in the DCE and introduced on a page directly before the first choice set. Table 1 gives an overview of these attributes and their corresponding levels.

Table 1: Attributes employed in the DCE

Attribute name	Alternative	Attribute level	Label
Minimum distance to residential areas		300 / 600 / **900** / 1600 / 2500	Distance
Size of the REF	Wind	small (5-10 turbines) / medium (18-25 turbines), large (35-50 turbines)	Size_W
	Solar	small (1-10 football f.) / medium (20-60 football f), large (100-150 football f)	Size_S
	Biomass	small (1-3 fermentation tanks) / medium (5-8 fermentation tanks), large (15-25 fermentation tanks) **The future SQ level is medium**.	Size_B
Number of REFs		1 / 2 / **3** / 4 / 5	Num_REFs
Protection of landscape view / share of landscape not used for renewable energy expansion (in %)		10 / 20 / **30** /40 / 50	Landscape
Long-distance transmission lines		**overhead** / underground	LDTL
Surcharge or rebate to your electricity bill in Euros per month (year)		-10 (-120) / -5(-60) / +2(24) / +7(84) / +14(168) / +23(276)	Cost

To allocate the attribute levels across choice sets, a Bayesian efficient design optimized for multinomial logit models (MNL) with labelled alternatives was generated. As an optimization criterion the C-error was used (Scarpa and Rose 2008). The prior values were taken from models estimated on data from the six focus groups as well as the pilot studies. The final design had 24 choice sets blocked into four groups of six choice sets. Respondents were randomly assigned to one of the four blocks. The order of appearance of the choice sets was randomized. Also, the order of the first three alternatives was randomized across respondents.

Respondents had the choice among four alternatives labelled as electricity from wind power, electricity from solar power, electricity from biomass, and “I do not care about the renewable energy technology”. Choosing the last option, a future status quo (FSQ) with a price of zero, would indicate that respondents do not to have influence on the type of renewable energy that could be installed in their 10 km surroundings, and that they accept the levels presented on this alternative. On each choice set respondents were asked to choose which alternative they prefer most (highlighted green in the second last row of the choice set), and which alternative they prefer least (last row of the choice set, highlighted red). On each choice set it was pointed out that participants should consider the area of 10 km surrounding their place of residence. If they lived in a large city, they should think of the landscape surrounding their city as this might be used for recreational purposes, for instance. Figure 1 gives an example of a choice set.

Figure 1: Example choice set

	Electricity from wind power	**Electricity from solar power**	**Electricity from biomass**	**No influence on renewable type**
Minimum distance to the edge of town	600m	2500m	300m	900m
Size of REFs	large (35-50 turbines)	large (15-25 fermentation tanks)	small (1-10 football fields)	medium
Number of REFs	4	5	5	3
Protection of landscape view	20%	50%	10%	30%
Long-distance transmission line	underground	Underground	overhead	overhead
Change of electricity bill per month (year)	+14€ (+168€)	-5€ (-60€)	+14 € (+168 €)	0 €
I choose …				
…. best option	□	□	□	□
…. worst option	□	□	□	□

5.2.2 Cut-off Elicitation

Prior to the DCE and before introducing the attributes, participants were asked to state their maximum (minimum) acceptable levels regarding four out of the six attributes. By asking the cut-off questions before and not after the DCE, I followed Swait's (2001) argument that thresholds should be based on individuals' past experience and not on information provided in the choice tasks. Similar to Aizaki et al. (2012), Ding et al. (2012), Moser and Raffaelli (2014), cut-off levels were presented on a card. Respondents were asked to select the category closest to their threshold. Attribute cut-offs were only requested from a randomized split sample of around 50% of the participants. For the attributes Distance and Num_REFs thresholds were elicited alternative-specific since it can be assumed that they depended on the type of renewable energy. Minimum requirements for Landscape and LDTL, however, were assumed to be generic. Thus, a total of eight cut-offs were elicited.

The wording of the Distance attribute cut-off question was as follows: “If you think of the expansion of renewable energies, which is the minimum distance renewable energy facilities should at least have to your place of residence?” Then, the different cut-off levels were presented on a card and respondents were requested to select the closest cut-off level to their thresholds (Table 2).

Table 2: Cut-off levels for attribute Distance

Renewable energy facilities to generate	***Minimum distance***					
	300m	600m	900m	1600m	2500m	I do not care about the minimum distance
electricity from						
wind power	o	o	o	o	o	o
Solar power	o	o	o	o	o	o
Biomass	o	o	o	o	o	o

The wording for the number of REFs and Landscape were similar. However, for the number of REFs respondents were asked to state their upper cut-off since utility is expected to be negatively influenced by this attribute. The cut-off question for the qualitative attribute LDTL was worded as follows: “Which statement applies to you the most?” The categories were: “A) When expanding the electricity grid, new long-distance transmission lines must be built overhead. B) When expanding the electricity grid, new long-distance transmission lines must be built underground. C) I do not care whether new transmission lines are built overhead or underground.” To reduce the added burden of collecting information on attribute cut-offs, no thresholds were elicited for the attribute Size and Cost. Other reasons for not requesting maximum acceptable levels with respect to the Cost attribute were that individuals were not familiar with the good under evaluation and that no detailed information about the choice sets had been given at this this stage of the survey.

5.3 Modelling Approach

The modelling approach is based on the random utility theory, with a utility function U for respondent n and alternative i in choice task t characterised by price p and non-price attributes $\boldsymbol{x}$ of the experimental design, and a random error term ε:

$$U_{nit} = -\alpha_n' p_{nit} + \boldsymbol{\beta}_n' \boldsymbol{x}_{nit} + \varepsilon_{nit} \quad , \qquad (1)$$

where α and β are parameters to be estimated and ε is assumed to be identically and independently distributed (*iid*) and related to the choice probability with a Gumbel distributed error term. Let the sequence of choices over T_n choice tasks for respondent n be y_n, i.e. y_n = $\langle i_{n1}, i_{n2}, \dots, i_{nT_n} \rangle$.

In a mixed logit (MXL) model, heterogeneity across respondents is introduced by allowing β_n to deviate from the population means following a random distribution. In a MXL model, the unconditional choice probability of respondent n's sequence of choices is the integral of the logit formula over all possible values of η_{ni} weighed by the density of η_{ni}:

$$\Pr(y_n|\alpha_n, \beta_n) = \int \prod_{t_1=1}^{T_n} \frac{\exp(-\alpha_n' p_{nit} + \beta_n' x_{nit})}{\sum_{j=1}^{J} \exp(-\alpha_n' p_{njt} + \beta_n' x_{njt})} f(\eta_{ni}|\Omega) d\eta_{ni}, \tag{2}$$

where $f(\eta_{ni}|\Omega)$ is the joint density of parameter vector for price and K non-price attributes [α_n, β_{n1}, β_{n2}, … , β_{nK}], η_{ni} is the vector comprised of the random parameters and Ω denotes the parameters of these distributions (e.g. the mean and variance). This integral does not have a closed form and thus requires approximation through simulation (Train 2009).

Several discrete choice models have been proposed to accommodate non-compensatory preferences in general (e.g. Elrod et al. 2004, Martínetz et al. 2009), and attribute cut-offs in particular (Swait 2001). In a two-stage sequential choice approach, proposed by Manski (1977), Manski and Sinha (1989) and Swiat and Ben-Akiva (1987), attribute cut-offs are reflected through a non-compensatory choice set formation model (first stage), and a second stage compensatory decision evaluating the choices of the screened alternatives (Swait 2001). This approach, however, does not explicitly use cut-off information from the decision maker. It is computationally intensive to estimate and do not allow for cutoff violations / penalties (Ding et al. 2012). The linear compensatory model by Swait (2001), in which cut-offs are “soft" and whose violation translates into penalties during the evaluation of alternatives incorporates cut-offs as a behavioral phenomenon in the evaluation stage of the choice process (Swait 2001).

Estimating a model with “soft” cut-offs requires adding a penalty function to the utility function, associating the cut-offs with the penalties. The utility function, without recognizing the sequence of repeated choices, would then have the following form:

$$U_{ni} = \sum_{i \in C} \delta_i \, U(X_i) + \sum_{i \in C} \sum_{k} \delta_i (w_i \, \lambda_{ik} + v_k \kappa_{ik})$$

where δ_i indicates which alternative out C available alternatives on the choice set was chosen, X_i is the k dimensional vector that describes the good, w_k is the marginal disutility of violating the lower cut-off for attribute k, v_k is the marginal disutility of violating the upper cut-off for attribute k, λ_{ik} is a cut-off constraint variable for the lower limit cut-offs constraint and κ_{ik} is a cutoff constraint for the upper limit cut-offs (Swait 2001, Bush et al. 2009).

5.4 Results

5.4.1 Descriptive Statistics of the Survey

In total, 12,833 panel members were invited to take part in the survey out of which 220 could not be admitted to the survey due to quota restrictions with respect to age and gender. Finally, 4,027 individuals participated in the survey and 3,400 completed the questionnaire. After data cleaning, 3,390 useable interviews remained and are used in subsequent analysis. Of the total number of participants 1694 respondents faced the cut-off questions as detailed in Section 2.2. Table 3 provides means and standard deviations of the socio-demographic characteristics as well as the place of residence of the respondents.

Mean age was measured to be 42.64 (42.80 or the cut-offs sample) years for the no cut-offs sample while 46% (45%) of the respondents were females. The average level of education, which was recoded as years of school and university attendance, was calculated to be 14.14 (14.03) years. 19% (21%) of the people participating in the survey lived in large cities, 18% (16%) in a suburban area of a large city. The percentage of respondents residing in medium size or small cities was 35% (33%) while 29% (31%) lived in villages.

Table 3 further indicates that assigning respondents randomly to a sample resulted in very similar values. For each variable, a two-sample t-tests of equal means was conducted. The null-hypothesis of equal means could not be rejected in any case. Compared to the German average the total sample, however, consists of a very large share of respondents with a university degree. This introduces a bias towards high education people. Since the aim of the paper is not to aggregate WTP estimates, I use this non-representative sample and note that qualification when interpreting the results.

Table 3: Socio-demographic characteristics of participants

Variable	Sample: no cut-offs Mean (SD)	Sample: cut-offs Mean (SD)
Number of Respondents	1696	1694
Age (years)	42.64 (14.20)	42.80 (14.00)
Gender (1 =female)	0.46 (0.50)	0.45 (0.50)
Education (years of schooling)	14.14 (3.56)	14.03 (3.57)
Place of residence		
- Large city	0.19 (0.39)	0.21 (0.40)
- Suburban area of a large city	0.18 (0.38)	0.16 (0.37)
- Medium size or small city	0.35 (0.48)	0.33 (0.47)
- Village	0.29 (0.45)	0.31 (0.46)

Note: SD =standard deviation

5.4.2 Comparison of Preferences Across Split Samples

The analysis on the influences of cut-off elicitation questions starts by comparing chosen alternatives between split samples. Note that each respondent faced six choice sets which resulted in a total of 10,176 (10,264) choices. As Table 4 indicates electricity from solar power is with 38.91% (40.48%) the most chosen alternative followed by wind power (28.58% / 26.67%) and electricity from biomass (21.67% / 21.44%). In only 10.87% (10.44%) of the choice sets participants opted for the future status quo (FSQ). As already mentioned, choosing the FSQ meant that individuals would not have any influence on the type of renewable energy which will be developed in their ten km surroundings. Table 4 also reveals that participants who were asked to state their attribute thresholds chose the solar alternative slightly more often the respondents who were not required to report their cut-offs. The opposite effect is observed for wind power, electricity from biomass and the FSQ. A chi-squared test, however, indicates that the null-hypothesis of equal frequencies cannot be rejected (p-value =0.19).

Table 4: Chosen alternatives across split samples

Alternative	Frequency (%): no cutoffs	Frequency (%): cutoffs
Wind	2,906 (28.56)	2,810 (26.67)
Solar	3,959 (38.91)	4,114 (40.48)
Biomass	2,205 (21.67)	2,179 (21.44)
FSQ	1,106 (10.87)	1,061 (10.44)
Total	10,176 (100)	10,164 (100)

Table 5 presents the estimation results of a MNL as well as an MXL model for both split samples. Since the MXL models clearly outperform the MNL models and the estimated standard deviations in the MXL models are highly significant, there is strong evidence of unobserved preference heterogeneity suggesting that attributes should be modelled as random parameters. Consequently, in what follows I focus on analyzing the results of the MXL models. While the alternative-specific constants (ASCs) are included as fixed parameters, all non-cost attributes are specified to follow a normal distribution while the negative of the cost attribute is assumed to be log-normally distributed.

As suggested by the ASCs, respondents prefer, independently of the attributes, solar power over the FSQ. They still prefer, on average, the FSQ over electricity from biomass and wind power. This effect, however, is only statistically significant in the sample where cut-offs were elicited. On average and independently of the treatment, participants prefer new REFs to be built further away from their place of residence. This effect is most pronounced for wind turbines followed by electricity from biomass and solar power. Compared to REFs with medium size, respondents are more likely to choose an alternative with small REFs (Size_Ws, Size_Ss, Size_Sb). For the sample with no cut-off elicitation this effect is, however, indistinguishable from zero with respect to electricity from solar and biomass. Unlike previous expectations, the number or REFs did not influence the probability of choosing an alternative. As suggested by the estimated standard deviation, which are shown in the lower part of Table 5, there is, regarding the number of REFs as well as the other attributes, a high degree of unobserved preference heterogeneity in the sample.

In line with prior expectations, respondents, on average, are more likely to choose an alternative with larger landscape shares not used for future renewable energy development. The sign and t-value of the attribute LDTL, which captures whether LDTL are built overhead or underground, indicates that respondents prefer new transmission lines to be built underground. Finally, the cost attribute has a negative sign and is highly significant showing that respondents are cost sensitive.

Table 5: Comparison of parameter estimates across split samples

Attributes	**MNL: no cutoffs**		**MNL: cutoffs**		**MXL: no cutoffs**		**MXL: cutoffs**	
	Coef.	|t-value|	Coef.	|t-value|	Coef.	|t-value|	Coef.	|t-value
Means								
ASC_W	0.5475	6.37	0.4614	5.23	0.0907	0.55	-0.4268	2.49
ASC_S	1.5892	19.08	1.5268	17.71	1.4789	9.82	1.2221	7.78
ASC_B	0.5477	6.08	0.4046	4.39	-0.2315	1.31	-0.8558	4.52
Distance_W	0.0339	11.02	0.0411	13.02	0.0431	5.99	0.0670	8.93
Distance_S	0.0117	3.65	0.0110	3.35	0.0158	(2.24)	0.0314	4.62
Distance_B	0.0225	6.95	0.0283	8.49	0.0184	2.29	0.0409	5.12
Size_Ws	0.3538	4.83	0.4132	5.55	0.2767	1.56	0.4207	2.32
Size_Ss	0.1458	2.43	0.2633	4.31	0.6109	5.03	0.6528	5.24
Size_Bs	0.1534	2.18	0.2486	3.41	0.2394	1.32	0.5608	3.59
Size_Wl	-0.3559	5.47	-0.2621	3.94	-0.3701	2.77	-0.3330	2.37
Size_Sl	-0.3898	6.47	-0.3957	6.50	-0.2936	2.31	-0.3773	2.85
Size_Bl	-0.3463	5.23	-0.3298	4.91	-0.4062	2.94	-0.3429	2.47
Num_REFs_W	0.0097	0.50	-0.0163	0.82	0.0204	0.46	0.0379	0.83
Num_REFs_S	-0.1098	6.16	-0.0806	4.40	-0.0714	1.81	-0.0269	0.66
Num_REFs_B	-0.0441	2.20	-0.0401	1.96	-0.0630	1.27	-0.0015	0.03
Landscape	0.0027	3.19	0.0112	12.80	0.0081	3.79	0.0316	12.76
LDTL	0.3279	12.31	0.3837	13.96	0.8200	11.53	0.9922	13.16
Cost	-0.0451	30.46	-0.0477	31.39	-2.5652	39.33	-2.5151	38.32

Attribute	**MNL: no cutoffs**		**MNL: cutoffs**		**MXL: no cutoffs**		**MXL: cutoffs**	
	Coef.	|t-value|	Coef.	|t-value|	Coef.	|t-value|	Coef.	|t-value
Standard deviations								
Distance_W					0.1156	11.48	0.1154	12.07
Distance_S					0.1323	12.96	0.0945	10.10
Distance_B					0.1246	11.13	0.1095	10.40
Size_Ws					2.0789	8.12	2.1683	7.90
Size_Ss					1.0375	4.10	1.2995	5.02
Size_Bs					1.4146	3.76	0.7635	2.26
Size_Wl					1.2210	4.93	1.5623	(6.85)
Size_Sl					1.1910	4.31	1.7884	8.26
Size_Bl					1.4236	5.78	1.4838	6.76
Num_REFs_W					0.7365	17.45	0.6329	15.48
Num_REFs_S					0.5821	15.35	0.6011	15.70
Num_REFs_B					0.5934	13.61	0.6040	13.31
Landscape					0.0395	11.08	0.0483	13.58
LDTL					1.2845	15.17	1.4016	15.15
Cost					1.2038	21.95	1.2091	21.45
Observations	10176		10164		10176		10164	
AIC	24368.4		23571.6		20850.0		20535.9	
BIC	24523.5		23726.6		21134.3		20820.2	
Log-likelihood	-12166.2		-11767.8		-10392.0		-10235.0	
Parameters	18		18		33		33	

Based on the results obtained from the MXL models, Table 6 depicts marginal WTP estimates for the sample in which cut-offs were elicited and the treatment where no attribute constraints were collected. Mean WTP, which is measured in EUR per month and household, was calculated by taking the ratio of marginal (dis)utilities of the non-cost attributes to the marginal disutility of the cost attribute.

For an 100m increase in the distance to wind power facilities, respondents are, on average, willing to pay 0.27 EUR (0.40) EUR per household and month. The corresponding figures for solar and biomass are 0.10 EUR (0.19 EUR) and 0.12 EUR (0.24), respectively. For a decrease in the size of the REF from medium to small, participants who were assigned to the treatment with cut-off elicitation are willing to pay 2.50 EUR (wind), 3.84 EUR (solar) and 3.34 EUR (biomass). Respondents of the treatment without threshold statements are only willing to pay for a decrease of the size of solar fields (3.89 EUR).

For an increase in the size of the REF participants require compensations which are calculated to be -2.33 EUR (-1.98 EUR) for electricity from wind power, -1.85 (-2.25 EUR) for solar power and -2.56 EUR (-2.04 EUR) for electricity from biomass. For an increase in the share of the landscape surrounding the respondents' place of residence which would not be used for renewable energy expansion, mean WTP amounts to 0.05 EUR (0.19 EUR). For LDTLs to be built underground rather than overhead, participants are willing to pay 5.17 EUR (5.91 EUR per month and household. Remarkably, mean WTP estimates for those attributes for which cut-offs were stated, namely Distance, Num_REFs, Landscape and LDTL, mean WTP is always higher. No systematic effect is observed for Size for which attribute thresholds were not collected.

To test whether these effects are statistically significant, a Poe et al. (2005) test was conducted. Results of this test are presented in the last column of Table 6. As suggested by the p-value of the test, mean WTP is indeed significantly higher for the attribute Landscape when thresholds were collected. Although only at the ten percent level of significance, similar effects are observed with respect to the distance of REFs to produce electricity from wind and biomass power. An explanation for these effects may be that explicitly asking participants on attribute constraints might have led to preference construction and preference learning. Since individuals might not have been familiar with the specific good, i.e. amenity impacts of different renewable energies, they might have constructed their preferences based on cut-off elicitation questions. Simon et al. (2004) observed that preference construction often takes place by constraint satisfaction.

In the next step, this might have induced respondents to accept higher levels of the attribute Cost in order not to violate their self-imposed attribute constraints. Remember that no attribute cut-offs were collected concerning Cost. Another fact which points into the same direction is that the attributes Landscape, Distance_W and Distance_B were those attributes for which the highest number of respondents indicated to have constraints. A detailed analysis on the sample with cut-off elicitation is conducted in the next section.

Table 6: Marginal WTP estimates across split samples

Attribute	MXL: no cutoffs		MXL: Cutoffs		Poe-test
	Mean WTP	95%-CI	Mean WTP	95%-CI	p-value
Distance_W (EUR / 100m)	0.27	0.18;0.37	0.40	0.30;0.49	0.09
Distance_S (EUR / 100m)	0.10	0.01;0.19	0.19	0.11;0.27	0.11
Distance_B (EUR / 100m)	0.12	0.02;0.22	0.24	0.15;0.34	0.07
Size_Ws (EUR)	n.s.		2.50	0.40;4.61	
Size_Ss (EUR)	3.85	2.34;5.35	3.89	2.42;5.36	0.61
Size_Bs (EUR)	n.s.		3.34	1.50;5.18	
Size_Wl (EUR)	-2.33	-4.00;-0.66	-1.98	-3.63;-0.33	0.33
Size_Sl (EUR)	-1.85	-3.44;-0.26	-2.25	-3.82;-0.67	0.58
Size_Bl (EUR)	-2.56	-4.26;-0.86	-2.04	-3.66;-0.42	0.26
Num_REFS_W (EUR / site)	n.s.		n.s.		
Num_REFS_S (EUR / site)	n.s.		n.s.		
Num_REFs_B (EUR / site)	n.s.		n.s.		
Landscape (EUR / %)	0.05	0.02;0.08	0.19	0.16;0.22	0.00
LDTL (EUR underground)	5.17	4.17;6.16	5.91	4.86;6.95	0.40

To complete the discussion on the influence of prior cut-off elicitation a Heteroskedastic Logit was estimated in which the scale parameter was specified as a function of whether cut-offs were elicited or not. Model results, which can be found in the appendix of this paper, show an increase in the scale parameter when cut-offs were required. Thus, the error variance in the sample with cut-offs elicitation was lower suggesting that choices are more consistent in this treatment again pointing to learning effects.

5.4.3 Cut-off Analysis

In this section, the sample in which attribute cut-offs were elicited is analyzed in detail. Table 7 starts by presenting frequencies of cut-off statements. As already mentioned, cut-offs were elicited for the alternative-specific attributes Distance and Num_REFs as well as the generic attributes Landscape and LDTL. This resulted in a total of eight cut-off values. No thresholds were collected for Size as well as Cost.

Table 7 reveals that a vast majority of respondents indeed reported minimum (maximum) requirements. The attribute with the highest number of participants indicating a threshold is the share of the landscape not used for future renewable energy development (around 90%) followed by the distance to wind farms (roughly 85%). At the other extreme, around 46% of participants did not state attribute constraints regarding LDTLs and the number of solar fields.

The highest number of respondents who stated their cut-off at the most severe level, also referred to as cut-off severity (Moser and Raffaelli 2014), is 54.13% for the attribute Landscape followed by the distance to REFs to produce electricity from biomass (40.73%). Compared to wind and biomass, it appears that cut-off frequency and severity is always lowest for electricity from solar power when looking at the alternative-specific attributes.

Table 7: Frequencies of cut-off statements

Cut-off	Frequency (%) Wind	Frequency (%) Solar	Frequency (%): Biomass	Frequency (%): Generic
Distance				
I do not care	251 (14.82)	653 (38.55)	272 (16.06)	
300m	83 (4.90)	341 (20.13)	69 (4.07)	
600m	138 (8.15)	180 (10.63)	133 (7.85)	
900m	296 (17.47)	212 (12.51)	232 (13.70)	
1600m	327 (19.30)	149 (8.80)	298 (17.59)	
2500m	599 (35.36)	159 (9.39)	690 (40.73)	
Num_REFs				
I do not care	575 (33.94)	775 (45.75)	556 (32.82)	
5	164 (9.68)	241 (14.23)	64 (3.78)	
4	88 (5.19)	114 (6.73)	39 (2.30)	
3	251 (14.82)	208 (12.28)	175 (10.33)	
2	263 (15.53)	172 (10.15)	265 (15.64)	
1	353 (20.84)	184 (10.86)	595 (35.12)	
Landscape (generic)				
I do not care				166 (9.80)
10%				53 (3.13)
20%				97 (5.73)
30%				256 (15.11)
40%				205 (12.10)
50%				917 (54.13)
LDTL				
I do not care				778 (45.95)
Overhead				34 (2.01)
Underground				881 (52.04)

Next, Table 8 reveals that many participants are willing to violate their self-stated cut-offs when having to make trade-offs against other attributes of renewable energy expansion. Along the sequence of six choice sets, in 39.24% (wind), 17.35% (solar) and 42.99% (biomass) of the choices people opted for an alternative where the distance level of the chosen alternative was lower than the self-stated minimum. The percentage of choices in which the number of REFs of the chosen alternative exceeded the attribute constraint was 29.73% (wind), 21.58% (solar) and 37.90% (biomass). Being the highest frequency of cut-off violation, 51.08% of the choices involved a violation of the attribute Landscape. The corresponding figure of the attribute LDTL is calculated to be 22.42%.

An explanation for choosing alternatives that violate self-stated cut-offs is that when considered in isolation the minimum (maximum) acceptable level of each attribute may actually reflect decision makers' purpose. However, when traded-off against other attributes, individuals may be willing to either change or violate cut-offs because the additional benefit is greater than the cost caused by the violation recognizing the opportunity cost of self-reported cut-offs (Swait 2001, Bush et al. 2009). Another reason of cut-off violation may be the design of the DCE, or, more specifically, the status quo alternative implemented in the present study. Although the cost of choosing the status quo alternative was zero, participants had still to accept the levels of the other attributes. For instance, the share of the landscape not used for future renewable energy development shown on the FSQ was always 30%. This figure is lower than the minimum requirement of roughly 2/3 of the respondents.

Table 8: Percentage of cut-off violation across the six choice sets

Attribute	Wind	Solar	Biomass	Generic
Distance	39.24	17.35	42.99	
Number of REFs	29.73	21.58	37.90	
Landscape				51.08
LDTL				22.42

Table 9 presents the results of a MXL in which cut-off violations are integrated by means of utility penalties following the Swait (2001) approach. Thus, in addition to the attribute parameters eight coefficients representing cut-off penalties are presented. Again, the ASCs are included as fixed parameters, all non-cost attributes as well as the utility penalties are specified to follow a normal distribution while the negative of the cost attribute is assumed to be log-normally distributed.

Compared to the model results presented in Table 5, the inclusion of cut-off parameters changes results quite substantially. First of all, all cut-off parameters turned out to be negative and highly significant. This result reveals that attribute thresholds play an important role when choosing among different renewable energy alternatives. At the same time, several of the attribute parameters become indistinguishable from zero when cut-off penalties are included. This applies to the distance to wind turbines and biogas power station as well as the share not used for future renewable energy development. This finding probably reflects the fact that these tree attributes were also those with the highest degree of cut-off severity (see also Table 7).

This non-significance of the main parameters can be interpreted such that respondents do not care about these attributes as long as their minimum requirements are met. Regarding the number of REFs, Table 9 shows some unexpected results. When cut-off parameters are considered, participants now experience a positive utility when number of REFSs to produce electricity from wind power and biogas is increased. However, utility becomes negative when the self-imposed cut-off is violated. Signs and levels of significance of the attributes concerning the size of the REF - cut-offs were not elicited for this attribute - are unaffected by the modeling approach. For the distance to REFs to generate electricity from solar power as well as the attribute concerning LDTS, the attribute parameter and the cut-off penalty turned out to be significant with the expected sign. The lower part of Table 9 further the standard deviations estimated for the main as well as the cut-off penalties. Sign, magnitude and the level of statistical significance clearly indicates that there is a large degree of unobserved preference heterogeneity across respondents. They further suggest cut-off penalties and thus whether decision are made based on cut-offs vary considerable across participants.

Finally, Table 9 clearly indicates that accounting for attribute improves model fit (see also Table 5).

Table 9: Accounting for attribute cut-offs using the Swait (2001) approach

Attribute	Attribute Parameter		Cut-off Parameter	
	Coef.	|t-value|	Coef.	|t-value|
Means				
ASC_W	2.3202	10.20		
ASC_S	2.7652	14.49		
ASC_B	1.3730	5.43		
Distance_W	0.01047	1.41	-0.1345	11.07
Distance_S	0.01851	3.16	-0.0642	4.60
Distance_B	0.0063	0.74	-0.0874	7.23
Num_REFs_W	0.1703	4.02	-0.3293	4.59
Num_REFs_S	-0.02111	0.56	-0.1084	1.61
Num_REFs_B	0.1302	2.55	-0.2321	3.29
Landscape	0.0044	1.17	-0.0329	6.87
LDTL	0.3048	4.05	-1.0675	10.07
Size_Ws	0.5240	2.91		
Size_Ss	0.7181	5.77		
Size_Bs	0.5770	3.60		
Size_Wl	-0.2880	2.15		
Size_Sl	-0.3502	2.36		
Size_Bl	-0.4321	3.11		
Cost	-2.6604	43.25		
Standard Deviaions				
Distance_w	0.0732	7.23	0.09438	6.57
Distance_S	0.0534	5.24	0.08601	3.45
Distance_B	0.0784	7.84	0.1000	6.58
Num_REF_W	0.3093	5.77	0.4296	3.02
Num_REFs_S	0.3269	7.47	0.1585	0.86
Num_REFs_b	0.2868	4.99	0.3909	3.45
Lanscape	0.0305	5.14	0.0398	5.89
LDTL	0.8313	6.86	1.0447	6.45
Size_Ws	2.1731	8.20		
Size_Ss	1.3811	5.84		
Size_Bs	0.7468	1.77		
Size_Wl	1.1849	4.84		
Size_Sl	1.6588	7.69		
Size_Lb	1.4540	6.51		
Cost	1.2837	23.11		
Observations	10164			
AIC	19952.6			
BIC	20322.9			
Log-likelihood	-9933.3			

5.5 Conclusions

This paper has analyzed the use of attribute thresholds (cut-offs) towards preferences for renewable energy development in Germany. Prior to the choice experiment, attribute cut-offs were stated for two alternative-specific and two generic attributes resulting in a total of eight cut-off values. In order to investigate the impact of prior cut-off elicitation on preferences, attribute constraints were only collected from half of the sample. The treatments, to which respondents were randomly assigned to, only differed in whether cut-offs were reported or not.

I have found that REFs to produce electricity from solar power is the most preferred alternative followed by electricity from biomass and wind power. Particularly with respect to wind farms respondents clearly prefer REFs to be constructed further away from their place of residence. Other attributes that influence public's preferences towards renewable energies are the size of the REF, the share of the landscape not used for future renewable energy expansion, whether new LDTLs are built overhead or underground as well as the influence on the electricity bill.

I have further found that prior cut-off elicitation tends to affect WTP estimates positively. For one of the attributes, the share of landscape not used for future renewable energy expansion, mean WTP has been calculated to be four times larger when thresholds were reported. An explanation for these results may be respondents' unfamiliarity with some of the attributes of renewable energy alternatives which might have led respondents to construct their preferences based on attribute constraints.

By further analyzing the split-sample with cut-off elicitation, I have found that a large majority of respondents appeared to have thresholds which were most severe for the attributes Landscape and the alternative-specific attributes with respect to electricity from wind and biomass power.

Nevertheless, many choices involved the violation of self-imposed thresholds. To account for this effect in the choice model, I have specified MXL model following the idea of soft cut-offs proposed by Swait (2001) capturing cut-off violations through negative impacts on the utility function, i.e. utility penalties. I have found that including cut-off penalties improves the fit of the model.

Furthermore, many of the penalty parameters have turned out to be significant indicating that cut-offs and their violations are relevant in the choice process. This indicates that respondents employed threshold-based non-compensatory decision strategies.

These results points to the need of further investigating the use of attribute cut-offs in environmental valuation. To the best of my knowledge this was only the second attempt to study threshold-based decision strategies in an environmental context. Further research might use other cut-off elicitation formats. I have implemented a card showing different levels for each attribute. An alternative approach could be to ask open-ended questions.

Furthermore, I have followed the Swait (2001) model assuming attribute cut-offs to be exogenous to the choice process. As pointed out by Ding et al. (2012) or Moser and Raffaelli (2014), cut-off endogeneity is potentially an important issue.

5.6 References

Aizaki, H., Sawada, M., Sato, K., Kikkawa, T. 2012. A noncompensatory choice experiment analysis of Japanese consumers' purchase preferences for beef. Applied Economics Letters 19(5), 439–444.

Álvarez-Farizo, B., Hanley, N. 2002. Using conjoint analysis to quantify public preferences over the environmental impacts of wind farms. An example from Spain. Energy Policy 30(2), 107–116.

Bermann, A., Colombo, S., Hanley, N. 2008. Rural versus urban preferences for renewable energy developments. Ecological Economics 65(3), 616–625.

BMWi 2017. Aktuelle Informationen: Erneuerbare Energien im Jahr 2016. Bundesministerium für Wirtschaft und Energie: Berlin. https://www.bmwi.de/Redaktion/DE/Artikel/Energie/arbeitsgruppe-erneuerbare-energien-statistik-02.html (last access: 14.11.2018)

Bush, G., Colombo, S., Hanley, N. 2009. Should all Choices Count? Using the Cut-Offs Approach to Edit Responses in a Choice Experiment. Journal of Environmental and Resource Economics 44(3), 397– 414.

Cantillo, V., Amaya, J., Ortúzar, J. de D. 2010. Thresholds and indifference in stated choice surveys. Transportation Research Part B: Methodogical 44(6), 753–763.

Cicia, G., Cembalo, L., del Giudice, T., Palladino, A. 2012. Fossil energy versus nuclear, wind, solar and agricultural biomass: Insights from an Italian national survey. Energy Policy 42, 59–66.

Danielis, R., Marcucci, E. 2007. Attribute cut-offs in freight service selection. Transportation Research Part E: Logistics and Transportation Review 43(5), 506–515.

Dimitropoulos, A., Kontroleon, A. 2009. Assessing the determinants of local acceptability of wind-farm investment: A choice experiment in the Greek Aegean Islands. Energy Policy 38(5), 1842–1854.

Ding, Y., Veeman, M.M., Adamowicz, W.L. 2012. The influence of attribute cut-offs and consumers' choices of a functional food. European Review of Agricultural Economics 39(5), 745–769.

Elrod, T., Johnson, R.D., White, J. 2004. A new integrated model of noncompensatory and compensatory decision strategies. Organizational Behavior and Human Decision Processes 95(1), 1–19.

Ek, K. 2005. Public and private attitudes towards “green” electricity: the case of Swedish wind power. Energy Policy 33, 1677–1689.

Hensher, D.A. 2006. How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics 21, 861–878.

Hensher, D., Rose, J. 2012. The Influence of Alternative Acceptability, Attribute Thresholds and Choice Response Certainty on Automobile Purchase Preferences. Journal of Transport Economics and Policy 46(3), 451–468.

Huber, J., Klein N.K. 1991. Adapting Cutoffs to the Choice Environment: The Effects of Attribute Correlation and Reliability. Journal of Consumer Research 18(3), 346–357.

Ju, H.C., Yoo, S.H. 2014. The environmental cost of overhead power transmission lines: the case of Korea. Journal of Environmental Planning and Management 57(6), 812–828.

Kosenius, A.-K., Ollikainen, M. 2013. Valuation of environmental and societal trade-offs of renewable energy sources. Energy Policy 62, 1148–1156.

Krueger, A.D., Parsons, G.R., Firestone, J. 2011. Valuing the Visual Disamenity of Offshore Wind Power Projects at Varying Distances from the Shore: An Application on the Delaware Shoreline. Land Economics 87(2), 268–283.

Ladenburg, J. 2009. Visual impact assessment of offshore wind farms and prior experience. Applied Energy 86(3), 380–387.

Ladenburg, J., Lutzeyer, S. 2012. The economics of visual disamenity reductions of offshore wind farms—Review and suggestions from an emerging field. Renewable and Sustainable Energy Reviews 16, 6793–6802.

Leong, W., Hensher, D.A. 2012. Embedding decision heuristics in discrete choice models: a review. Transport Reviews: A Transnational Transdisciplinary Journal 32(3), 313–331.

Marcucci, D., Gatta, W. 2011. Regional airport choice: Consumer behaviour and policy implications. Journal of Transport Geography 19(1), 70–84.

Martínez, F., Aguila, F., Hurtubia, R. 2009. The constrained multinomial logit model: a semi-compensatory choice model. Transportation Research Part B: Methodological 43(3), 365–377.

McNair, B.J., Bennett, J., Hensher, D.A., Rose, J.M. 2011. Households' willingness to pay for overhead-to-underground conversion of electricity distribution networks. Energy Policy 39(5), 2560–2567.

Mentzakis, E., Stefanowska, P., Hurley,, J. 2011. A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. Health Economics, Policy and Law 6(3), 405–433.

Moser, R., Raffaelli, R. 2014. Does attribute cut-off elicitation affect choice consistency? Contrasting hypothetical and real-money choice experiments. Journal of Choice Modelling 11, 16–29.

Moser, R., Raffaelli, R. 2012. Consumer preferences for sustainable production methods in apple urchasing behaviour: a non-hypothetical choice experiment. International Journal of Consumer Studies 36(2), 141–148.

Peschel, A.O., Grebitus, C., Colson, G., Hu, W. 2016. Explaining the use of attribute cut-off values in decision making by means of involvement. Journal of Behavioral and Experimental Economics 65, 58–66.

Román, C., Arencibia, A. I., Feo-Valero, M. 2017. A latent class model with attribute cut-offs to analyze modal choice for freight transport. Transportation Research Part A: Policy and Practice 102, 212-227.

Scarpa, R., Rose, J. 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. The Australian Journal of Agricultural and Resource Economics 52, 253–282.

Schroeder, A., Oei, P.Y., Sanders, A., Hankel, L. 2013. The integration of renewable energies into the German transmission grid—A scenario comparison. Energy Policy, 61, 140–150.

Simon, D., Krawczyk, D.C., Holyoak, K.J. 2004. Construction of Preferences by Constraint Satisfaction. Psychological Science 15, 331–336.

Simon, J.A. 1995. A behavioral model of rational choice. Quarterly Journal of Economics 69 (1), 99–118.

Strazzera, E., Mura, M., Contu, D. 2011. Combining choice experiments with psychometric scales to assess the social acceptability of wind energy projects: A latent class approach. Energy Policy 48, 334–347.

Swait, J. 2001. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodigical 35(10), 903–928.

Swait, J., Ben-Akiva, M. 1987. Incorporating random constraints in discrete choice models of choice set generation. Transportation Research Part B: Methodogical 21 (2), 91–102.

Train, K. 2009. Discrete choice methods with simulation. Cambridge University Press: Cambridge.

Tversky, A. 1972. Elimination by aspects: A theory of choice. Psychological Review 79 (4), 281–299.

5.7 Appendix

Table A1: Accounting for scale differences across split samples

Attribute	Coef.	|t-value|
ASC_w	0.4724	8.08
ASC_S	1.4713	24.52
ASC_B	0.4506	7.37
Distance_W	0.03549	16.74
Distance_S	0.01052	4.86
Distance_B	0.02402	10.87
Size_Ws	0.3636	7.36
Size_Ss	0.1982	4.91
Size_Bs	0.1904	3.98
Size_Wl	-0.2895	6.57
Size_Sl	-0.3657	9.00
Size_Bl	-0.3197	7.16
Num_REFs_W	-0.003047	0.23
Num_REFs_S	-0.09054	7.47
Num_REFs_B	-0.03970	2.94
Lanscape	0.006684	11.63
LDTL	0.3351	18.10
Cost	-0.04368	37.78
Scale		
Cut-offs elicited	0.1113	4.62
Observations	20340	
AIC	47969.6	
BIC	48146.4	
Log-likelihood	23965.8	

6 Stated Preferences towards Renewable Energy Alternatives in Germany - do the Consequentiality of the Survey and Trust in Institutions Matter?

Malte Oehlmann[a], Jürgen Meyerhoff[a]

Abstract

This research concerns the effect of consequentiality and trust in institutions on willingness to pay estimates towards the expansion of renewable energy in Germany. We use four information treatments which differ in terms of the information participants received prior to a discrete choice experiment. Treatments differ with respect to a consequentiality device and the institution which would be responsible for providing the good under evaluation. After finishing the choice tasks, respondents stated their perceived consequentiality and trust in institutions. We find perceived policy consequentiality to be strongly associated with the trust individuals have in both providing institutions. Moreover, compared to the treatments which did not highlight the consequences of the survey, participants are more inclined to perceive their responses to be at least somewhat consequential when the consequentiality device was presented. However, willingness to pay estimates do neither differ across treatments nor by the level of perceived consequentiality. We speculate that as the expansion of renewable energy is strongly debated with the public having a wide range of beliefs and political views, the requirements for consequential choices are not met.

Keywords: Consequentiality; Discrete choice experiment, Trust in institutions; Externalities of Renewable Energies

Published in: Journal of Environmental Economics and Policy (2017), 6 (1): 1-16, http://dx.doi.org/10.1080/21606544.2016.1139468

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of Environmental Economics and Policy on 29 Jan 2016, available online: https://www.tandfonline.com/doi/full/10.1080/21606544.2016.1139468

[a] Technische Universität Berlin, Institute for Landscape Architecture and Environmental Planning, Straße des 17. Juni 145, D-10623 Berlin, Germany

6.1 Introduction

Whether respondents truthfully reveal their preferences in stated preference surveys is subject to a debate that is on-going since stated preference methods were introduced. Carson and Groves (2007) set out two conditions for stated preference surveys to provide useful information about individuals' preferences. First, the agent needs to care about what the outcomes of those actions might be. Second, the agent answering the survey needs to view his or her responses as potentially influencing decision-makers. Surveys satisfying these criteria are viewed as 'consequential', which, in turn, is a crucial condition to achieve incentive compatibility (Carson and Groves 2007). Since inconsequential surveys might explain the hypothetical bias often reported in the literature (Vossler, Doyon, and Rondeau 2012; Carson, Groves, and List 2014), consequentiality also affects the external validity of stated preference surveys (Vossler and Watson 2013). The theoretical framework of incentive compatibility and consequentiality of stated preference questions has largely been laid out by Carson and Groves (2007, 2011), Vossler, Doyon, and Rondeau (2012) and Carson, Groves, and List (2014).

This research is complemented by several studies investigating different aspects of consequentiality empirically (e.g. Bulte et al. 2005; Herriges et al. 2010, Vossler, Doyon, and Rondeau 2012; Interis and Petrolia 2014). The overall finding of the empirical literature is that the probability to accept an offered bid decreases (1) when participants are informed that survey results would be made available to decision-makers and (2) when they state that there is at least some change of impacting policy. We contribute to the body of empirical research by studying consequentiality in a discrete choice experiment (DCE) on the expansion of renewable energies in Germany. Within the survey, we implement two different approaches. First, we exogenously vary the information participants receive prior to the DCE. Second, we use follow-up questions asking participants whether they believe that their responses will be considered by policy-makers.

A closely related issue to the consequentiality of the survey is the trust in the institution which provides the good under evaluation. Oh and Hong (2012) showed that trust in governments and willingness to pay (WTP) estimates are positively associated. Therefore, we also investigate the influence of two different providing institutions on stated choices.

We use an approach with four treatments that differed in terms of the information participants were provided with prior to the DCE. A consequentiality device was shown to half of the sample highlighting that the results of the survey would be made available to decision-makers. The information treatments also differ with respect to the institutional context, i.e. whether the

National Government or the 16 State Governments of Germany will be responsible for implementing the policy. After finishing the choice tasks, all respondents were asked to state their perceived consequentiality as well as their trust in the two different institutions on a four-point Likert scale. Perceived consequentiality was measured using two items. First, respondents were asked whether they believe that their choices affect policy (policy consequentiality) and second, whether they think that they will have to pay for the good evaluated (payment consequentiality).

The main objective of this research is to investigate whether WTP estimates are (a) sensitive to the information people receive in the different treatments and (b) the beliefs about the consequentiality of the survey respondents stated after the DCE. In particular, we follow the 'knife-edge' argument by Carson and Groves (2007) predicting that respondents who perceive that there is some positive probability of their responses affecting actual policy behave similarly to each other because

they face the same incentive structure (Carson and Groves 2007, Interis and Petrolia 2014). We particularly focus on WTP estimates since they are the main measure of interest for decision-makers. In addition, we shed light on the determinants that influence perceived consequentiality especially with respect to the institutional context.

The main contribution of this research is that it simultaneously accounts for the effects of consequentiality and the institutional context. Moreover, we are only aware of one study (Vossler and Watson 2013) investigating the drivers which impact perceived consequentiality. We further contribute to the literature on public preferences towards renewable energy development. While the Energiewende in Germany, which aims at a comprehensive transition of the energy system towards renewable energies, is in general broadly supported by the public (BMWi 2014), the construction of renewable energy facilities (REFs) often faces substantial opposition at the local level. The DCE was, therefore, intended to quantify the external costs form the development of renewable energies and high-voltage transmission lines (HVTL) in the 10 km surroundings of the respondents' place of residence.

In the remainder of the paper, the following section reviews the empirical literature relevant to this research and outlines our expectations. Section 3, then, describes the study design before the econometric approach is briefly presented in Section 4. Section 5 details the results, and Section 6 discusses the main findings.

6.2 Empirical Literature and Expectations

Three strands of empirical literature are of particular relevance to our research: a) studies investigating the effect of consequentiality devices to inform participants about the consequences of their responses, b) research on the influences of perceived consequentiality on stated choices, and c) studies concerning the role of different providing institutions. Within the first line of literature, Bulte et al (2005) used a split sample approach with one treatment informing about the potential consequentiality, i.e. that the survey results will be made available to policy makers and could thus serve as a guide for future decisions. Their application was a CV study to value different policies to protect seal populations. The consequentiality device was found to lead to significantly smaller WTP estimates compared to the inconsequential treatments. Similar results were presented by Landry and List (2007) who compared choices among three hypothetical treatments (fully hypothetical, 'cheap talk script', consequentiality device") to a treatment with real economic consequences. They found that the consequential and 'cheap talk' treatments were statistically indistinguishable from real responses. The application of this study was a CV to value sports memoralia.

Within the second strand of literature, which uses subjective measures of consequentiality, Nepal et al. (2009) applied a series of follow-up questions to a CV scenario to construct different indices of consequentiality. Across these indices they found, in contrast to theoretical predictions, statistical evidence for a positive association between consequentiality and WTP towards efforts to mitigate global climate change. The first to examine perceived consequentiality in a DCE setting were probably Vossler et al. (2012). They employed four treatments in their study; three of them used real payments. The fourth one was labelled the stated preference treatment as no project could be implemented as a direct result of the participants' choices. Each choice set offered a binary choice between a tree plantation project (yes vote) and the status quo (no vote). They found the WTP functions for the three real payment treatments to be statistically identical, but the WTP function for the stated preference treatment to be statistically different from all real payment functions. However, when they conditioned their analysis on the belief that responses had more than a weak impact on policy, stated and real WTP functions were statistically identical. This observation is the 'knife-edge' result predicted by Carson and Groves (2007).

In 2013, Vossler and Watson compared survey responses from verified voters with the outcome of a parallel public referendum on conservation and preservation programs. In their survey they recorded respondents' perceptions regarding the potential policy impact (consequentiality).

They found a negative hypothetical bias, i.e., those who did not believe that the results were consequential responded less often 'yes' to the referendum questions. The WTP estimates for this group were also lower. Interis and Petrolia (2014) used a split sample approach with a binary- and a multinomial-choice setting. In an application to value coastal restoration programs in Louisiana, they fail to observe the 'knife-edge' result in the binary data set, but do observe it in the multinomial data.

Among those who analysed the influence of a consequentiality device as well as perceived consequentiality were Herriges et al. (2010). They found support for the equality of WTP distributions among those believing that the survey is at least minimally consequential while those who believed that the survey is not consequential had statistically different distributions. Moreover, the consequentiality device was shown to be positively associated with the perceived degree of consequentiality. Herriges et al. (2010) implemented a split sample approach using a dichotomous choice CV referendum for water quality programs in Iowa Lakes. In one sample, they did not only inform people that survey results would be provided to decision makers, but also gave respondents the information that the responsible agency is already using information from the survey for decision making and would continue to use it.

Regarding the influence of trust on stated WTP we were only able to identify a very limited number of studies. Oh and Hong (2012) developed a theoretical model on the link between citizens' trust in government and their WTP. They concluded that a positive association existed between both trust and WTP and therefore public projects could be hindered by prevailing distrust toward governments. In one of the few empirical studies Remoundou et al. (2012) examined whether preferences as well as WTP estimates were sensitive to the trust respondents have in the providing institution. Less trust in an institution can, as they argued, result in less valid results since the consequentiality of the survey results is seen as weaker in that case. Using a split sample approach, the authors presented two different managing institutions: the European Commission versus the National Government of Greece. Although they found significant differences in trust towards the two institutions, both coefficients and WTP estimates were, however, not statistically different across their treatments.

From this literature we derive the following four expectations:

- Expectation I: Perceived consequentiality significantly differs across treatments; in treatments with consequentiality device a higher degree of perceived consequentiality is expected.

- Expectation II: Compared to participants who were not shown the consequentiality device, WTP estimates are expected to be significantly lower for those who were assigned to a treatment explicitly informing respondents about the use of the survey results.
- Expectation III: WTP estimates are expected to be higher when respondents are assigned to a treatment with the providing institution being the State Government. The level of trust is expected to be higher for the State Government compared to the National Government.
- Expectation IV: Compared to respondents who state that their choices will definitely not be taking into account, WTP estimates are expected to be lower for those who state that their responses will at least partially impact policy. This is the 'knife-edge' reported in the literature.

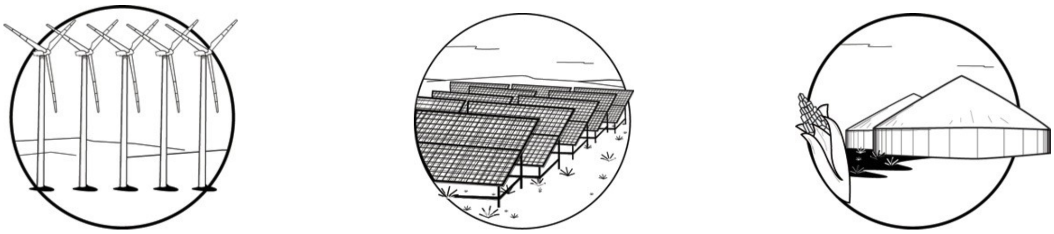
6.3 Survey Design

6.3.1 Study Design and Survey Administration

As a pre-study to elaborate the questionnaire, six focus groups in different towns spread over Germany were conducted in October 2012. Based on the feedback received during the focus groups the DCE was revised and finally tested in two pilot studies: one with colleagues and the second with members of the survey companies' online panel.

The final version of the questionnaire started by requesting socio-demographic variables with respect to age, gender and education. Then, it was highlighted that the survey aimed at three renewable energy sources: onshore wind energy, solar energy (photovoltaic power stations installed on open fields) and biomass (biogas power stations). Afterwards, respondents were asked whether they had REFs as well as HVTLs in the ten km surroundings of their place of residence. If they lived in a large city, they were pleased to take into account the area surrounding their city. Ladenburg (2014), for instance, showed that preferences towards renewable energies are dependent on the experience people have with the different REFs. In our survey prior experience was elicited by the categories "yes", "no" or "I do not know". In order to give some visual guidance, the question was each time accompanied by a pictogram (see Figure 1).

Figure 1: Pictograms presented to respondents



Wind energy refers to electricity generation with single wind turbines and wind farms exclusively on the mainland.	Solar energy refers exclusively to the production of electricity with photovoltaic systems in the open landscape, so-called solar fields.	Biomass refers to the production of biogas and its electricity and includes both the biogas plant as well as the cultivation of the required biomass (such as corn)

The bottom line reports the text used to define the renewables in the survey

The DCE was introduced as follows: “Renewable energies as well as the electricity grid will be expanded in Germany. On the following choice sets you can choose among different alternatives of renewable energy development. Please think of renewable energy facilities to be built in the ten km surroundings of your place of residence. If you live in a large city please consider the surrounding area of your city. You can choose among the following alternatives:

- Electricity from wind energy (wind farms)
- Electricity from solar energy (solar fields)
- Electricity from biomass (biogas power stations)
- I do not care about the type of renewable energy generation (you will not have any influence on the type of renewable energy which will be developed in the ten km surroundings of your place of residence.)”

Each alternative was again visualised with the pictograms shown in Figure 1. The last alternative mentioned, a future status quo, was used instead of a commonly employed status quo alternative. It had a price of zero. Choosing this alternative indicates that respondents, compared to the other available alternatives, do not care about the type of renewable energy that would be developed in their surroundings. At the same time they agree with the levels of this alternative (see Figure 2) that were the same on all choice sets.

Figure 2: Example choice set

	Electricity from wind energy	Electricity from solar energy	Electricity from Biogas	Don’t care about the type of Renewable Energy
Minimum distance to the edge of town	600 m	2500 m	300 m	900 m
Size of REFs	large (35-50 turbines)	large (15-25 fermentation tanks)	small (1-10 football fields)	medium
Number of REFs	4	5	5	3
Protection of landscape view	20%	50%	10%	30%
High-voltage transmission line	underground	underground	overhead	overhead
Change in electricity bill per month (year)	+14 € (+168 €)	-5 € (-60 €)	+14 € (+168 €)	0 €
I choose …				
…. best option	□	□	□	□
…. worst option	□	□	□	□

The choice attributes, which were described to respondents after the preamble of the DCE are reported in Table 1. They relate to the minimum distance of REFs to the edge of town, the size and number of REFs, whether new HVTLs are built overhead or underground, and the change of the electricity bill (positive and negative). Additionally, the attribute “protection of landscape view” relates to the connected share of the landscape in the ten km surroundings which will not be used for renewable energy development in the future. Figure 2 gives an example of a choice set.

On each choice set respondents were requested to choose which alternative they would prefer within a ten km surroundings of their place of residence (highlighted green in the second last row of the choice set), and which alternative they prefer least (last row of the choice set, highlighted red). As the choice refers to the ten km surroundings, respondents living in large cities were asked to think of the landscape around that city assuming that they might use it for recreational purposes.

Table 1: Attributes and their levels

Attribute name	Alternative	Attribute level	Label
Minimum distance to residential areas in (m)		300 m, 600 m, **900 m**, 1600 m, 2500 m	Distance
Size of REF	Wind	small (5-10 turbines), **medium (18-25 turbines)**, large (35-50 turbines)	Size of REFs
	Solar	small (1-10 football fields), **medium (20-60 football fields)**, large (100-150 football fields)	
	Biomass	small (1-3 fermentation tanks), **medium (5-8 fermentation tanks)**, large (15-25 fermentation tanks)	
Number of REFs		1, 2, **3**, 4, 5	Number of REFs
Protection of landscape view (%)		10, 20, **30**, 40, 50	Landscape
HVTL		**overhead,** underground	HVTL
Cost in € per month (year) and household: surcharge or rebate to electricity bill		-10 (-120), -5(-60), +2(24), +7(84), +14(168), +23(276)	Cost

Note: Levels of the future status quo-alternative are written in bold; HVTL = high-voltage transmission line; REF = renewable energy facility

In order to combine the attribute levels to choice sets, we generated a Bayesian efficient design optimised for Multinomial Logit (MNL) models with labelled alternatives using Ngene software. As optimisation criterion we used the C-error which aims at minimising the variance of the sum of the marginal WTP estimates (Scarpa and Rose 2008). To allow for uncertainty in the value of the prior, modified latin hypercube sampling was applied and 1000 draws were taken for each parameter prior from uniform distribution. The prior values were obtained from models estimated based on data from the focus groups and the pilot studies. The resulting design had 24 choice sets that were blocked into four blocks with each six sets. The order of appearance was randomised. Also the order of the first three labelled alternatives was randomised across respondents, so it was the same for each respondent.

In addition, the questionnaire comprised several questions with respect to attitudes, acceptances and fairness aspects. At the end of the survey further socio-demographic variables such as the place of residence, the household's income as well as the interest in the topic of renewable energy development were recorded. Interest was measured on a four-point scale from 'not at all' to 'strongly'.

6.3.2 Treatments

In order to investigate effects of the consequentiality devices as well as the institutional context, we created four information treatments (Table 2). Across the treatments we varied the level of consequentiality, i.e. whether it was mentioned that the information is provided to decision makers, and the institution that is mainly responsible for the transformation of the energy system towards renewable energies. The treatment-specific information page was shown to respondents after the choice sets were introduced and right before the first choice set was presented. The corresponding information was printed in bold. Note that the four treatments only differed with respect to the information participants received prior to the DCE, and that respondents were randomly assigned to one of the four treatments.

Table 2: Information treatments

Treatment label	Framing
NatCon(T1)	*Institution: National Government* / consequentiality device: yes *'Before we present the choice sets to you we would like to point out that the National Government and the National Ministries are responsible for the expansion of renewable energies. The results of the study will be made available to the National Ministries. Thus, you and the other respondents can influence the expansion of renewable energies.'*
Nat(T2)	*Institution: National Government* / consequentiality device: no *'Before we present the choice sets to you we would like to point out that the National Government and the National Ministries are responsible for the expansion of renewable energies.'*
StaCon(T3)	*Institution: State* Government / consequentiality device: yes *'Before we present the choice sets to you we would like to point out that the government and the ministries of your State are responsible for the expansion of renewable energies. The results of the study will be made available to State Ministries. Thus, you and the other respondents can influence the expansion of renewable energies.'*
Sta(T4)	*Institution: State* *Government* / consequentiality device: no *'Before we present you the choice sets we would like to point out that the government and the ministries of your state are responsible for the expansion of renewable energies.'*

Note: Nat = National; Sta = State; Con = Consequentiality mentioned

6.3.3 Perceived Consequentiality and Trust in Institutions

After the sequence of six choice sets respondents were presented with two items on the perceived consequentiality of the survey: policy consequentiality and payment consequentiality. In both cases four-point response scale with no middle category were used. The wording of the two items was as follows:

- Policy consequentiality: 'To what extent do you believe that the choices among renewable energies you have just made will be taken into account in future decision making concerning the expansion of renewables energies? My choices will be:' Definitively considered / Rather considered / Rather not considered / Definitively not considered
- Payment consequentiality: 'To what extent do you believe that the choices among renewable energies you have just made will affect your future electricity bill? My choices will:' Definitively affect / Rather affect / Rather not affect / Definitively not affect.

Subsequently, and again on a four-point Likert scale, participants were asked to what extent they trust the two different institutions. Trust in this study refers to the extent respondents believe that the institutions will take into account the interest of the public during the transformation of the energy system. The institutions were the same as used for the design of the four treatments, thus the National Government and the governments of the 16 States. All respondents were asked to state their trust in both institutions regardless of the treatment they were assigned to. The wording of this question was as follows:

- Trust in institutions: "When do you think about whether the interests of the population will be taken into account during the process of renewable energy expansion: To what extent would you trust the following institutions?'

=> National Government and Ministries (Trust_Nat)

=> Government and ministries of my State (Trust_Sta).

Response scale: Completely trust / Rather trust / Rather not trust / Completely not trust

6.4 Econometric Approach

For our analysis of the stated choices we use McFadden's (1974) random utility model as point of departure. It assumes that the researcher does not possess complete information concerning the individual decision maker n; individual preferences are therefore the sum of a systematic (V) and an unobservable or stochastic component (ε), where V is an indirect utility function. If the stochastic component is distributed independently and identically (IID) and follows a Gumbel distribution, the conditional choice probability that alternative i is chosen is defined as:

$$P_{ni} = \frac{\exp(\mu\beta_k X_{ik})}{\sum_{j \in C} \exp(\mu\beta_k X_{jk})}, \qquad (1)$$

where the scale parameter (μ) of the error distribution is confounded with the parameter vector ßk and generally normalized to one, Xik is attribute k of alternative i. However, this model is for a couple of reasons not sufficient for analysing our data. One reason is that the survey was carried out across Germany and we thus expect that a significant degree of taste heterogeneity among respondents. While the conditional logit model can capture observed heterogeneity partly through interaction terms, given that the sources of heterogeneity are known, it is likely that a significant amount of unobserved heterogeneity will remain. A second reason is that during the interview each respondent faced six choices and thus our data is a type of panel. We decided therefore to estimate the widely used Mixed Logit (MXL) model (Train 2009). Unobserved taste heterogeneity is in our specification taken into account by assuming the non-cost attribute parameters to be from a normal distribution, and the cost attribute to be from a log-normal distribution. The MXL probability was simulated each time using 2000 Halton draws. All models were estimated in Stata using the mixlogit program provided by Hole (2007) and are based on the the alternatives respondents selected to be their best.

6.5 Results

6.5.1 Descriptive Statistics

The nation-wide online survey took place in September and October 2013. In total 12,833 members of an online panel were invited to take part. Out of these 220 could not be admitted to the survey as quota restrictions were already fulfilled (a quota system for age and gender was applied). Finally, 4,027 persons took part in the survey and 3,400 completed the questionnaire. After data checking, in particular with respect to item non-response, 3,213 usable interviews remained. According to the AAPOR-Standard Definitions (AAPOR 2009) this leads to a response rate of 25.94%.

With respect to each treatment, Table 3 provides several socio-demographic variables as well as respondents' prior experiences with REFs and their interest in the topic of renewable energy development. While the mean values of all socio-demographic characteristics do not differ much across treatments, it is at the same time visible that the complete sample comprises a large share of respondents with a university degree. This indicates that the sample is biased towards higher education levels, a problem which has previously been reported in other studies using online surveys (e.g. Kosenius and Ollikainen 2013, Meyerhoff et al. 2012). The same issue is also pointed out by Lindhjem and Navrud (2011) in their comparison of Internet panels to other survey modes. However, here the share of participants with a high levels of education is even for an online sample high. A reason for this is probably the topic of the survey. However, as we do not aim at aggregating WTP estimates, this issue is of less concern for the present paper.

Table 3 also depicts statistics with respect to respondents' place of residence. With slightly higher values for large cities, each time about one third of participants stated to live in villages / rural areas, small cities and large cities. Next, people's experiences with REFs is reported. Note that "I do not know"-responses were coded as if they had no REFs in their surroundings. Overall, participants seem to be highly affected by renewable energy development. In all treatments, more than 65% stated to have wind energy facilities in the ten km surroundings of their place of residence. More than half of the respondents indicated that they have solar facilities, while around 45% have biogas power stations. The percentage of people with HVTLs in their vicinity was measured to be around 88%. Participants also indicated to be highly interest in the topic of renewable energy expansion. The mean values, which are above 3 in all four sub-groups, indicate that respondents are highly interested in this topic.

Overall, the sampled populations of the four treatments are very similar to each other. We conducted a series of two-sample t-tests of equal means. The only null-hypothesis we had to reject was with respect of equal proportions of women in Nat(T2) compared to StaCon(T3) and Sta(T4).

As a consequence, possible biases in WTP estimates across the four treatments will not result from the variables shown in Table 3.

Table 3: Socio-demographic characteristics across treatments

Variable	NatCon (T1) Mean (SD)	Nat (T2) Mean (SD)	StaCon (T3) Mean (SD)	STA (T4) Mean (SD)
Age (years)	42.93 (13.87)	42.53 (14.54)	42.38 (13.89)	42.64 (13.91)
Female respondents	0.45 (0.50)	0.42 (0.49)	0.47 (0.50)	0.47 (0.50)
Education				
Middle-school incomplete	0.07 (0.25)	0.06 (0.25)	0.05 (0.22)	0.06 (0.25)
Middle-school degree	0.23 (0.42)	0.25 (0.43)	0.25 (0.43)	0.24 (0.43)
High-school degree	0.28 (0.45)	0.27 (0.44)	0.26 (0.44)	0.26 (0.44)
University degree	0.42 (0.49)	0.41 (0.49)	0.43 (0.50)	0.44 (0.50)
Income (€ / month)	3052.24 (1595.31)	3080.22 (1569.31)	3056.14 (1431.46)	3019.97 (1582.14)
Place of residence				
Large city	0.36 (0.48)	0.38 (0.49)	0.35 (0.48)	0.37 (0.48)
Small city	0.32 (0.47)	0.32 (0.47)	0.35 (0.48)	0.33 (0.47)
Village	0.33 (0.47)	0.29 (0.46)	0.30 (0.46)	0.30 (0.46)
Experience with REFs				
Wind	0.68 (0,47)	0.65 (0.48)	0.68 =0.47)	0.66 (0.47)
Solar	0.55 (0.50)	0.51 (0.50)	0.52 (0.50)	0.52 (0.50)
Biogas	0.46 (0.50)	0.47 (0.50)	0.47 (0.50)	0.42 (0.50)
HVTLs	0.88 (0.33)	0.90 (0.30)	0.89 (0.32)	0.89 (0.32)
Interest in the topic renewable energies	3.13 (0.66)	3.11 (0.71)	3.08 (0.67)	3.09 (0.65)
Respondents	788	779	817	829

Note: SD = Standard deviation;

Next, Table 4 shows how much time respondents spent on the survey page informing about consequentiality and the providing institution. Time spent on a webpage might not be an ideal indicator of whether respondents read instructions, but it can serve as a proxy. Table 4 shows that respondents spent more time on the survey page when the consequentiality device was presented. A Wilcoxon rank-sum test indicates that the median values for NatCon(T1) and StaCon(T3) are significantly higher than those for Nat(T2) and Sta(T4). This could be interpreted as an indication that respondents assigned to a treatment with consequentiality device read the additional information.

Table 4: Average time respondents spent on treatment survey page in seconds

	Trimmed mean	Median
NatCon(T1)	11.01	10
Nat(T2)	7.45	7
StaCon(T3)	11.18	11
Sta(T4)	7.85	7

Note: For the trimmed mean 1% of the lowest and highest values were excluded

Table 5 reports the answers on the perceived policy consequentiality of the survey. For the complete sample, 90.45% of the respondents do think that there is at least some probability that their responses will be taken into account by policy makers (definitely yes, rather yes, rather no). With respect to each treatment, this figure is lowest in Nat(T2) (88.19%) and highest for StaCon(T3) (91.92%).

Table 5: Perceived policy consequentiality across treatments in %

Category	NatCon(T1)	Nat(T2)	StaCon(T3)	Sta(T4)
Definitely no	8.88	11.81	8.08	9.53
Rather no	63.83	59.82	64.26	64.54
Rather yes	22.22	24.13	22.52	21.95
Definitely yes	4.06	4.24	5.14	3.98

Note that, as mentioned earlier, we also asked respondents on their beliefs about influences of their choices on the electricity bill. As illustrated in a cross-comparison in Table 6, we find that participants stated a much higher level of payment consequentiality compared to policy consequentiality. For instance, 140 participants stated to think that their choices would definitely affect policy making (definitely yes) while 1,329 participants believe their choices to impact their future electricity bill. This finding is not plausible since policy consequentiality would entail payment consequentiality. We suspect that participants did not relate their answers concerning payment consequentiality to the choices they made during the DCE, but rather to general changes in the electricity price, which has significantly increased in Germany over the last years. Thus, we subsequently only consider responses to policy consequentiality.

Table 6: Perceived policy and payment consequentiality

	Definitely no	Rather no	Rather yes	Defnietely yes	Total
Definetely no	89	32	16	3	140
Rather no	327	287	120	3	737
Rather yes	775	642	573	37	2027
Definetely yes	138	43	65	61	307
Total	1329	1004	774	104	3221

Note: Due to missing values with respect to the item on perceived payment consequentiality the sample size is reduced to 3,211 respondents here

Table 7 associates the trust in both the National Government and the State Governments. It is easily visible that respondents have significantly more trust in the Governments of their State than in the National Government when it comes to taking peoples interests and preferences into account. The frequency of those who responded that they definitively trust the National Government or their State Government does not differ much. However, 1459 respondents (45%) stated that they would fully or rather trust their State Government while only 813 respondents (25.3%) stated that they would fully or rather trust the National Government. Looking at the endpoints of the response scales, only 31 respondents fully trust both institutions while 384 respondents do not have any trust in the two institutions when it comes to taking their preferences into account. That trust in institutions differs significantly is also supported by a chi-square test of equality in frequency distributions.

Table 7: Stated trust in institutions

		National Government (Trust_Nat)				
		Fully	Rather	Rather not	Not at all	Total
State govern. (Trust_Sta)	Fully	31	34	18	9	92
	Rather	7	490	756	114	1,367
	Rather not	16	201	930	137	1,284
	Not at all	4	30	52	384	470
Total		58	755	1,858	644	3,213

Note: A Chi-squared test of equal levels of trust across institutions is rejected at the 1% level of significance

6.5.2 Determinants of Perceived Policy Consequentiality

This section investigates the factors that are expected to possibly impact perceived policy consequentiality. Table 8 reports the results of a binary logit with dependent variable taking a value of 0 if the participant stated that his or her responses would definitely not be taken into account and 1 otherwise.

Consistent with Expectation II and Section 5.4 we thereby follow the ”knife-edge” argument.

Table 8 shows a very strong association between perceived policy consequentiality and trust in institutions. Particularly, respondents who report high levels of trust in the National Government are more likely to state that their responses will be taken into account by policy makers. The effect is less pronounced for trust in the State Government. One reason for this finding might be that the National Government takes a key role in the transformation process of the energy system. Participants who do not trust the National Government are thus less likely to believe that their responses are considered.

Interestingly, Table 8 also indicates that the information treatments respondents received prior to the DCE have a significant influence on consequentiality perception. Compared to treatment Nat(T2), which did not highlight the consequences of the survey, participants are more inclined to perceive their responses to be at least somewhat consequential if they were assigned to a treatment with consequentiality device. The magnitude of this effect is slightly larger if the institutional context is the State Government.

Similar to Vossler and Watson (2013), our model on perceived policy consequentiality shows a rather weak link between consequentiality perception and socio-demographic covariates. Our estimates only show significant effects regarding the variables ‘age” finding that older respondents are observed to be less likely to state that their responses have some policy impacts. For the variable “interest”, we observe that the higher the interest in the expansion of renewable energies, the lower the probability of believing that their choices will be consequential with any positive probability.

Table 8: Binary Logit Model on Perceived Consequentiality

Variable	Coefficient	|t-value|
Trust_Nat	1.3122	10.03
Trust_Sta	0.4315	4.38
NatCon(T1)	-0.3573	1.98
StaCon(T3)	-0.4735	2.59
Sta(T4)	-0.2628	1.50
Age	-0.0133	2.63
Gender (female)	0.0172	0.13
Middle-school degree	0.1217	0.45
High-school degree	0.1002	0.37
University degree	0.0574	0.22
Income	0.0403	0.94
Place of residence (base =large city)		
Small city	0.2826	1.76
Village	0.1541	0.88
Experience with REFs		
Wind	-0.2497	1.83
Solar	-0.0243	0.18
Biogas	0.0508	0.35
HVTLs	0.3393	1.53
Interest	-0.4799	5.20
Constant	-5.9457	9.82
Observations	3213	
Pseudo r-squared	0.168	
AIC	1723.5	
Log-likelihood (constant)	-1012.7	
Log-likelihood (model)	-842.7	
Parameter	19	

6.5.3 The Influence of the Information Treatments on WTP estimates

The last section has shown that the information given in the treatments has some influence on subjective measures of consequentiality. Now, we investigate whether this is also the case concerning WTP estimates towards the expansion of renewable energies in Germany. The following table reports the results from MXL models estimated for each treatment. The alternative-specific constants (ASC) each time show that, all else held constant, respondents prefer solar power over wind power and wind power over electricity from biogas in the ten km surroundings of their place of residence. With the exception of NatCon(T1), electricity from biogas is, on average, not preferred over the future status quo alternative. As indicated by the large standard deviations, which are highly significant for all ASCs, there is a high degree of unobserved preference heterogeneity. With respect to the attributes, the likelihood of choosing an alternative increases with the distance REFs have to respondents' place of residence. By contrast, this probability decreases with the size of the REF. No effect is found for the number of REFs in peoples' surroundings. However, at least in treatment NatCon(T1) and StaCon(T3) significant standard deviations indicate some degree of taste heterogeneity with respect to this attribute. The coefficients for the attribute 'landscape' are positively significant indicating that larger shares in the landscape not used for future renewable energy development are associated with a utility gain. Participants also prefer HVTLs to be built underground rather than overhead. Finally, the 'cost' attribute has a negative sign and is highly significant. Thus, the more expensive an alternative is the less likely are respondents to choose it.

Table 9: Mixed Logit Models by Treatment

Variable	**NatCon(T1)**		**Nat(T2)**		**StaCon(T3)**		**Sta(T4)**	
	Coefficient	|t-value|	Coefficient	|t-value|	Coefficient	|t-value|	Coefficient	|t-value|
Means								
ASC (wind)	0.7370	5.03	0.4502	3.55	0.6968	4.77	0.7839	5.71
ASC (solar)	1.7909	12.04	1.4284	11.37	1.6417	11.03	1.3973	10.14
ASC (biogas)	-0.4332	2.54	-0.1979	1.33	0.0859	0.55	-0.1967	1.25
Distance	0.0525	9.79	0.0473	10.22	0.0501	9.50	0.0472	10.02
Size of REFs (small)	0.4889	4.57	0.3976	3.96	0.4149	3.92	0.3717	3.60
Size of REFs (large)	-0.4759	4.47	-0.4806	4.88	-0.4697	4.50	-0.4514	4.37
Number of REFs	0.0482	1.48	0.0164	0.55	0.0521	1.56	0.0137	0.45
Landscape	0.0169	5.66	0.0163	5.84	0.0175	5.81	0.0192	6.66
HVTLs (underground)	0.7280	8.41	0.7099	9.21	0.8101	9.29	0.7620	8.89
Cost	-2.7179	28.49	-2.5979	32.10	-2.6715	29.03	-2.6519	30.99
Standard deviations								
ASC (wind)	2.6002	15.49	2.0744	14.96	2.5826	(15.59)	2.4908	16.06
ASC (solar)	2.5119	16.09	2.0645	16.11	2.5739	16.43	2.4474	16.83
ASC (biogas)	2.4915	3.27)	2.1099	13.76	2.2832	13.40	2.3534	(13.79)
Distance	-0.0727	9.43	-0.0517	6.72	-0.0753	9.61	-0.0558	7.25
Size of REFs (small)	1.1305	6.92	1.1043	7.82	1.1936	8.09	-1.2052	8.68
Size of REFs (large)	-0.9331	5.26	-0.9788	6.38	1.0013	6.00	1.1409	7.17
Number of REFs	0.1747	2.08	-0.1238	1.22	0.1884	2.09	0.0662	0.56
Landscape	-0.0391	8.34	-0.0371	9.14	-0.0417	8.97	-0.0410	9.68
HVTLs (underground)	1.1700	10.05	0.9290	8.39	1.1823	9.97	1.2797	11.85
Cost	1.3369	15.75	1.1634	14.56	1.2981	14.13	1.1869	15.68
Observations	18912		18696		19608		19896	
Pseudo r-squared	0.171		0.136		0.159		0.158	
AIC	9229.7		9492.7		9735.9		10039.9	
Log-likelihood (constant)	-5543.4		-5471.9		-5763.4		-5934.7	
Log-likekihood (model)	-4594.8		-4726.4		-4847.9		-4999.9	
Parameters	20		20		20		20	

Note: REF = renewable energy facility; AIC = Akaike Information Criterion; ASC = alternative-specific constant

Table 10 presents the mean marginal WTP estimates as well as the 95-confidence intervals derived from the four mixed logit models. They are expressed in € per month and household. Taking the example of NatCon(T1), respondents, on average, are willing to pay €0.33 for the REF to be built 100 m further away from the edge of town. For a change in the size of the REF from middle to small, participants would pay €3.00. For a change from middle to small, however, compensations amounting to €2.95 are required. For one percent increase in the share of the landscape in the ten km surroundings of the respondent's place of residence, which would not be used to build REFs, the WTP is calculated to be €0.10. Respondents also have a significant WTP for building future HVTLs underground instead of overhead. It is estimated to be €4.51 in NatCon(T1). Comparing mean marginal WTPs across treatments, it appears that all confidence intervals overlap indicating that there are no systematic differences across split samples. Moreover, a series of Poe et al. (2005) tests on pair-wise WTP differences between treatments confirm that mean WTP estimates do not differ by treatment.

Table 10: Mean marginal WTP estimates by treatment in € per month

Attribute	**NatCon(T1)**	**Nat(T2)**	**staCon(T3)**	**Sta(T4)**
	WTP CI	WTP CI	WTP CI	WTP CI
Distance (100m)	0.33 (0.24 / 0.41)	0.32 (0.25 / 0.40)	0.31 (0.23 / 0.39)	0.33 (0.25 / 0.41)
Size of REFs (small)	3.00 (1.67 / 4.40)	2.72 (1.34 / 4.09)	2.58 (1.23 / 3.94)	2.61 (1.17 / 4.04)
Size of REFs (large)	-2.95 (-4.36 / -1.54)	-3.28 (-4.70 / -1.87)	-2.92 (-4.32 / -1.53)	-3.16 (-4.66 / -1.67)
Number of REFs	n.s.	n.s.	n.s.	n.s.
Landscape (%)	0.10 (0.06 / 0.14)	0.11 (0.07 / 0.15)	0.11 (0.07 / 0.15)	0.13 (0.09 / 0.18)
HVTLs (underground)	4.51 (3.21 / 5.82)	4.85 (3.61 / 8.09)	5.04 (3.68 / 6.40)	5.34 (3.95 / 6.73)

Note: REF = renewable energy facility; HVTL = high-voltage transmission line; WTP = willingness to pay; CI = 95%-confidence interval

6.5.4 The Influence of Perceived Consequentiality on WTP estimates

So far we have analysed whether the four treatments affect WTP estimates. Now, we compare mean marginal WTP estimates among those who at least partially believe in the consequentiality of their choices (positive consequentiality) and those who stated that they definitively do not think that their answers to the choice sets will affect policy (zero consequentiality). Table 11 presents the mean marginal WTP estimates for both groups derived from mixed logit models specified in the same way as the previous models. At a first glance it is already visible that WTP estimates do not differ systematically between sub-samples. While the mean WTPs for 'Distance' and 'Number of REFs' are larger for respondents who think that their choices have a zero probability of impacting policy, the estimates of 'Landscape' and 'HVTL' are larger for the other group. Overlapping confidence intervals, however, suggest that the equality of WTP distributions between both sub-samples cannot be rejected. This result is confirmed by again applying the Poe et al. (2005) test. The last column of Table 11 shows the p-values of the null-hypothesis that the mean WTP estimates of the sample "zero consequentiality" are larger than those of the sample "positive consequentiality" (Expectation IV). As a consequence, we do not observe the 'knife-edge' result in our data.

Table 11: Mean marginal WTP estimates by Policy Consequentiality Perception in € per month

	Zero consequentiality WTP (CI)	Positive consequentiality WTP (CI)	Poe eta al. (2005) test P-value
	N = 307	N = 2906	
Distance	0.35 (0.22 / 0.48)	0.32 (0.28 / 0.36)	0.07
Size of REFs (small)	2.86 (0.66 / 5.07)	2.81 (2.08 / 3.54)	0.31
Size of REFs (large)	-2.73 (-4.72 / -0.73)	-3.08 (-3.84 / -2.32)	0.57
Number of REFs	n.s.	0.22 (0.01 / 0.44)	-
Landscape	0.11 (0.05 / 0.17)	0.12 (0.10 / 0.14)	0.40
HVTLs (underground)	3.96 (2.04 / 5.87)	5.08 (4.39 / 7-78)	0.59

Note: REF = renewable energy facility; HVTL = high-voltage transmission line; WTP = willingness to pay; CI = 95%-confidence interval

6.6 Discussion and Conclusions

The main objective of this paper was to investigate whether WTP estimates are sensitive to (a) exogenously presented information on the institution that provides the good under valuation as well as whether survey results are made available to decision makers, and (b) the consequentiality perceptions of the participants. At the same time, this research contributes to the literature on preferences towards renewable energies and HVTLs. We find that, on average, participants prefer electricity from solar energy over wind energy, and electricity from wind energy over electricity from biogas in their 10 km surroundings. In contrast, Significant WTP estimates indicate that participants experience negative externalities from REFs and HVTLs in their 10 km surroundings. In particular, participants are willing to pay significant amounts to increase the distance to REFs as well as for HVTLs to be built underground rather than overhead.

Regarding our four information treatments, we find that WTP estimates do not differ significantly across treatments. From this, we conclude that neither the consequentiality device nor the institutional context influence stated choices. This result is contrary to our expectations concerning the influence of the information treatments (see Expectations II and III), but in line with Remoundou, Kountouris, and Koundouri (2012) who also found no association between the institutional frame and stated WTP. The consequentiality device, however, has been observed to impact WTP estimates in other studies (Bulte et al. 2005; Landry and List 2007). One explanation for this could be that respondents have not recognised the information treatment in our survey. We do not have any measure to confirm or reject this. The amount of time participants spent on the treatment page, however, suggests that the additional information was read. It also has to be noted that all respondents were told on the first screen of the questionnaire that the objective of the survey was to find out what people in Germany think about the expansion of renewable energies, and that the results would be used to inform decision-makers. The reason to do so was to motivate participants to proceed with the questionnaire. However, as this information was presented on the first page of the survey, it is likely that it was not salient anymore when respondents started to answer the sequence of choice tasks.

Noteworthy, we find the information treatments to have some impact on consequentiality perceptions when we pool those respondents who believe their choices to have at least some change of influencing policy. This result is in line with our Expectation I and with Herriges et al. (2010), who found a positive relationship between presenting a consequentiality device and respondents' beliefs about the policy impacts of the survey. Nevertheless, the impact of other factors such as respondents' interest in the topic, and, particularly, trust in institutions are observed to have much stronger effects on perceived consequentiality. This result emphasises the limited ability of the researcher to promote consequentiality as perceived consequences of the survey might be driven by many factors (similar findings were presented by Vossler and Watson 2013). Trust in institutions or respondents' interest, for instance, are beyond the design of a stated preference survey.

Another important result of our study is that different perceptions of policy consequentiality do not translate into varying WTP estimates. We followed the 'knife-edge' prediction by Carson and Groves (2007) and estimated models for individuals who at least partially perceive their choices as consequential and those who state that their choices will definitely not be taken into account by policy makers. We fail to observe the 'knife-edge' result, which contradicts our Expectation IV and findings presented by Nepal, Berrens, and Bohara (2009), Herriges et al. (2010), or Vossler, Doyon, and Rondeau (2012).

Overall, we do not find evidence in support for three out of four expectations. We only find weak evidence for a link between presenting a consequentiality device and perceived consequentiality.

The reasons why our findings differ to previously presented results are likely to be twofold. First, since we implement a series of multinomial choice questions, our study differs with respect to the choice format from previous research concerning consequentiality. While we implemented a choice set with four labelled alternatives, the literature has so far only used choice sets with two options. The only exception is Interis and Petrolia (2014) who found behavioural differences between a binary and a multinomial choice task when taking perceived consequentiality into account. This points to the need to further investigate the incentive structure of DCEs with more than two alternatives and a series of choice tasks. The incentive compatibility of such choice formats has yet to be shown.

Second, another important issue which distinguishes our study from previous research on consequentiality of stated preference questions is the good to be valued. For instance, Bulte et al. (2005) focused on declines in the seal population in the Waddensee (an estuary in the north of the Netherlands), while Herriges et al. (2010) aimed at water quality in primary recreational lakes of Iowa. We assume that these projects were not subject to very broad societal debate with many stakeholders involved and lobbies trying to influence public opinion. Unlike these studies, the transformation of the energy system including the expansion of renewable energy in Germany, which is the focus of this study, is a strongly debated topic with the public having a wide range of beliefs and political views. Vossler, Doyon, and Rondeau (2012) pointed out that requirements for consequential choices might not be present in such a context. They emphasise that when the good being valued is hotly debated with participants having entrenched political views or questioning the objectives of policymakers, beliefs about the strategic value of non-truthful voting may be stronger. Therefore, another topic for future research is to investigate the links between the nature of the environmental change to be valued, the trust that people have in the providing institutions, and the requirements of consequentiality.

Acknowledgements

The authors would like to thank two anonymous reviewers as well as Geesche M. Dobers for valuable comments and Julian Sagebiel for helping us conducting the Poe-test. Funding for this research was provided by the Federal Ministry of Education and Research in Germany and is gratefully acknowledged (Project: Efficient and fair allocation of renewable energy production at the national level, Fkz.01LA1110A).

6.7 References

AAPOR, 2009. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 6th Edition.

Bulte, E, Gerking, S., List, J. A., de Zeeuw., A, 2005. The effect of varying the causes of environmental problems on stated WTP values: evidence from a field study. *Journal of Environmental Economics and Management* 49, 330–342.

Camperon, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Carson, R.T., Groves, T., List, J.A., 2014. Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice. *Journal of the Association of Environmental and Resource Economists* 1, 171–207.

Carson, R.T., Groves, T., 2011. Incentive and information properties of preference questions: commentary and extensions. I*n:* Bennett, J., (ed.) The International Handbook on Non-Market Environmental Valuation. Cheltenham.

Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions. *Journal of Environmental and Resource Economics* 37, 181–210.

Federal Ministry of Economic Affair and Energy (BMWi), 2014. The Energy of the Future. First 'Energy Transition' Progress Report. Berlin.

Herriges, J., Kling, C., Liu, C.-C., Tobias, J., 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management* 59, 67–81.

Hole, A. R., 2007. Fitting mixed logit models by using maximum simulated likelihood. *Stata Journal* 7, 388-401.

Interis, M., Petrolia, D., 2014. The Effects of Consequentiality in Binary- and Multinomial-Choice Surveys. *Journal of Agricultural and Resource Economics* 39, 201-216.

Kosenius, A.-K., Ollikainen, M., 2013. Valuation of environmental and societal trade-offs of renewable energy sources. *Energy Policy* 62, 1148–1156.

Ladenburg, J., 2014. Dynamic properties of the preferences for renewable energy sources - a wind power experience-based approach. Energy 76, 542–551.

Ladenburg, J., Lutzeyer, S., 2012. The economics of visual disamenity reductions of offshore wind farms—Review and suggestions from an emerging field. Renewable and Sustainable Energy Reviews 16, 6793–6802.

Landry, C.E., List, J.A., 2007. Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field. *American Journal of Agricultural Economics* 89, 200–429.

Lindhjem, H., Navrud, S., 2011. Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. *International Review of Environmental and Resource Economics* 5, 309-351.

Meyerhoff, J., Angeli, D., Hartje, V., 2012. Valuing the benefits of implementing a national strategy on biological diversity - the case of Germany. *Environmental Science & Policy* 23, 109-119.

Nepal, M., Berrens, R.P., Bohara, A.K., 2009. Assessing Perceived Consequentiality: Evidence from a Contingent Valuation Survey on Global Climate Change. *International Journal of Ecological Economics and Statistics* 14, 14-29.

Oh, H., Hong, J. H., 2012. Citizens' trust in government and their willingness-to-pay. *Economics Letters* 115, 345-347.

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87, 353-365.

Scarpa, R., Rose, J. M., 2008. Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics* 52, 253-282.

Remoundou, K., Kountouris, Y., Koundouri, P., 2012. Is the value of an environmental public good sensitive to the providing institution? *Resource and Energy Economics* 34, 381-395.

Rose, J.M., Black, I.R., 2006. Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters* 17, 295–310.

Train, K., 2003. Discrete choice methods with simulation. Cambridge.

Vossler, C.A., Doyon, M., Rondeau, D., 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. *American Economic Journal: Microeconomics* 4, 145–171.

Vossler, C.A., Watson, S.B., 2013. Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization* 86, 137-147.

7 Synthesis

Passed years have seen growing interest in the application of discrete choice experiments (DCEs) to value environmental goods and services. However, the use of this method is still highly controversial (e.g. Haab et al. 2013, Hausman et al. 2012, Kling et al. 2012). The critique of stated preference (SP) methods is often associated with the possibility of deriving biased estimates due to the hypothetical nature of DCEs or, more generally, due to deviations from the assumptions of "rational" consumer theory underlying discrete choice analyses. In the DCE context, individuals' decisions that are not fully consistent with one more of these assumptions have been referred to as (behavioral) response anomalies (Johnston et al. 2017). As DCEs are implemented in surveys, the central question is whether such anomalies may be triggered by the design of the SP survey.

This dissertation aimed at analyzing response anomalies in DCEs for environmental valuation. It contributes to behavioral research in DCEs for environmental valuation within the broader concepts of construct and criterion validity.

The research has been structured along five major research questions elaborated in Chapter 1:

- Research question 1: How does choice task complexity influence survey drop-out rates, choice consistency and derived preferences?
- Research question 2: How does choice task complexity affect attribute non-attendance?
- Research question 3: How does choice task complexity influence the status quo effect and subsequently derived preferences?
- Research question 4: How do attribute cut-offs affect preferences?
- Research question 5: How does survey consequentiality influence preferences?

To answer these research questions, two specially tailored, large-scale, nation-wide online surveys were designed and evaluated. The first survey elicited preferences towards land-use changes in Germany. It employed a Design of Desings (DoD) approach originally introduced by Hensher (2004) in transportation research. This design plan enabled one to vary choice task complexity in a systematic manner by means of the design dimensionality, i.e. the number of choice sets, the number of alternatives, the number of attributes, the number of attribute levels and the level range. To the best of my knowledge this was the first time a DoD approach was implemented and its effects studied in an environmental context. In addition to the choice data, supplementary survey questions and paradata on survey drop-outs were used to address research questions 1 to 3.

The second survey aimed at eliciting preferences towards renewable energy development in Germany. Auxiliary survey questions were used to address research questions 4 and 5 thereby contributing to research on survey consequentiality and the use of attribute cut-offs, a topic that has so far received almost no attention in an environmental context.

The next section summarizes the main findings of the core chapters of this dissertation before presenting a discussion of these results and drawing overall conclusions. Next, some implications for researchers and practitioners are summarized. Finally, future research is highlighted.

7.1 Summary of Results

In this section the results of the five research articles are summarized. Besides answering the five research questions, this section also points to findings beyond the scope of the main research interests.

Using the dataset on preferences towards land-use changes in Germany, the first part of this thesis – Chapter 2, 3 and 4 – was concerned with the influence of choice task complexity on different behavioral outcomes addressed in research question 1, 2 and 3. Chapter 2 served as a point of departure for further analyzing impacts of task complexity. It was found that the probability to abandon the survey significantly increased with the number of choice sets and the number of attributes. Furthermore, designs with five alternatives showed a higher drop-out rate compared to those with only three alternatives. All other design dimensions did not significantly influence the probability to quit the survey.

With respect to influences of the design dimensions on choice consistency, the null hypothesis of no relationship between the design dimensionality and the error variance was each time rejected. However, some of these effects were not very pronounced (the number of alternatives, the number of choice sets). In addition, two interaction effects impacted the error variance significantly:

- The position of the choice set in the sequence of choices and the number of alternatives: Designs with four alternatives were observed to have higher levels of choice consistency at the beginning of the sequence, but within the course of that sequence error variance significantly increases.
- Number of attributes and attribute levels: A low number of attributes or levels had a negative effect on the error variance that increases and turned into a positive effect when the number of attribute levels increases.

With respect to welfare estimates, accounting for differences in the error variance due to the design dimensionality did not significantly affect willingness to pay (WTP) estimates.

Next, the influence of the design dimensionality on the attendance to attributes was investigated. Here, the most important result is that the relationship between attribute non-attendance (ANA) and the design dimensionality was not very strong. Nevertheless, results from using respondents' statements on the attendance to attributes as well as the inferred approach indicate that the number of choice sets partially (in each approach with respect to one out of three attributes analyzed) influenced ANA. The main driver behind ANA, however, was observed to be the number of alternatives. No influences were found with respect to the number of attributes, the number of attribute levels and the level range.

Next, Chapter 4 analyzed the influence of task complexity on the status quo effect. It was found that participants were more likely to choose the status quo alternative, the later a choice task was presented in the sequence of tasks. Also, the more alternatives a choice task comprised, the less likely respondents were to choose the status quo. No effects were observed with respect to the number of attributes and attribute levels. In contrast, participants were less inclined to choose the current situation when the level range was less extreme. In addition to the design dimensionality, a structural complexity measure (entropy) was used to explain status quo choices. Entropy is a single measure of choice task complexity and captures effects resulting from the similarity between alternatives in a choice task (Swait and Adamowicz 2001). It was found that the likelihood of staying at the status quo increased with entropy.

Chapter 4 also revisited the influence of the design dimensions on the error variance. Unlike Chapter 2 the model this time also accounted for unobserved preference heterogeneity. It was found that the variance of the error term decreased with an increasing number of choice sets. Choice consistency was further promoted when level ranges are narrow, but decreased when the level range broadened. The other design dimensions, including the number of alternatives, did not significantly impact choice consistency. In addition, Chapter 4 indicates that differences in the design dimensions significantly influence marginal as well as non-marginal welfare estimates. Calculations from two designs that caused the lowest and the highest number of status quo choices resulted in statistically different welfare estimates. This was true for marginal (mean WTP) as well as non-marginal (compensating variation) estimates. By including or excluding the parameter of the alternative-specific constant of the status quo in the calculations of the non-marginal welfare measures, even a change in the sign was observed. Thus, individuals experienced utility from the valued land use changes according to one design, but would experience a loss of utility from the same land use changes according to the other design.

Table 1 summarizes the main findings from Chapter 2, 3 and 4 with respect to each individual design dimension.

Table 1: Influence of design dimensions on drop-out rates, choice consistency, attribute attendance and status quo choices

Design Dimension	Drop-out rate	Choice consistency	Attribute attendance	Status quo choices
Number of choice sets	Increase	Quadratic*	Decrease	Increase
Number of alternatives	Increase for 5 alternatives	Decrease for 4 alternatives*	Decrease	Increase
Number of attributes	Increase	Decrease**	No effect	No effect
Number of attribute levels	No effect	Decrease**	No effect	No effect
Level range	No effect	Decrease	No effect	Increase

Note: * = Significant interaction between dummy 4 alternatives and number of choice sets; ** = Significant interaction between number of attributes and number of attribute levels

Findings of Chapter 2, 3 and 4 which go beyond the scope of the main research questions include that survey drop-out rates, choice consistency as well as status quo choices were also impacted by the socio-demographic characters age, gender and the level of education of the respondent. Furthermore, Chapter 4 revealed that the higher respondents perceived the quality of the landscape in their surroundings (stated status quo), the more likely they were to choose the status quo alternative. With respect to the main attribute parameters it was found that all coefficients were highly significant and had the expected sign. On average respondents preferred higher shares of forests in their surroundings, higher levels of biodiversity and less land conversion. Also, the coefficient for cost was negative and significant indicating that respondents were less likely to choose an alternative with a higher price.

Research question 4 and 5 were addressed by using data obtained through a survey on the development of renewable energies in Germany. Chapter 5, which studied the use of attribute thresholds as stated by participants prior to the DCE, found that prior cut-off elicitation impacted preferences. For one of the attributes, mean WTP was calculated to be four times larger when respondents were asked to state their minimum (maximum) thresholds before choosing among the renewable energy alternatives. Cut-off elicitation was further observed to have a negative impact on the error variance. By analyzing the split-sample with cut-off elicitation, it was found that a large majority of respondents appeared to have thresholds. Nevertheless, many choices in the DCE involved the violation of self-imposed thresholds. To account for this, a discrete choice model that allowed for cut-off violation by introducing utility penalties was implemented. Many of the estimated penalty parameters turned out to be significant.

Finally, Chapter 6 addressed the influence of survey consequentiality on preferences. Using a split sample approach, around halve of the respondents were presented a consequentiality device which informed them that the results of the survey would be made available to policy makers. After the DCE respondents were asked to state their perception of the consequences of their choices.

It was found that WTP estimates did not differ significantly across the information treatments. Thus, informing respondents about the consequences of their answers did not influence choices. However, the information treatments were found to have some impact on consequentiality perceptions when those respondents who believed their choices to have at least some change of influencing policy were pooled. Another important result of Chapter 6 was that different perceptions of consequentiality did not translate into varying WTP estimates.

Besides answering research question 5, Chapter 6 also found that the institutional context (state versus national government and ministries) did not influence choices. Moreover, perceived consequentiality was influenced by trust in instructions and the socio-demographic characteristics of the respondents.

Chapter 5 and 6 further revealed that, on average, renewable energy facilities to produce electricity from solar power was the most preferred alternative followed by electricity from biomass and wind power. Particularly with respect to wind farms, respondents clearly preferred renewable energy facilities to be constructed further away from their place of residence. Other attributes that influenced public's preferences towards renewable energies were the size of the facility, the share of the landscape not used for future renewable energy expansion, whether new long-distance transmission lines are built overhead or underground as well as the influence on the electricity bill.

Further insights on the efficient and equitable development of renewable energy plants in Germany, which are partially based on the results of the same survey, may be found in Drechsler et al. (2017).

7.2 Conclusions and Discussion

While the last section summarized the results from each individual research article, this section is intended to draw overall conclusions and discussing these results.

As presented in Section 1.2.2, response anomalies have been investigated in at least three streams of research: information procession limitations, systematic patterns of respondents' decision making and the incentive structure of DCEs. Information processing limitations in DCEs are often analyzed within the boundaries of passive bounded rationality (DeShazo and Fermo 2004). Here, the variance of the error term of the utility function serves as an indicator for choice consistency which is, dependent on the influence studied, typically interpreted in terms of fatigue (or boredom) on the one hand and learning or preference matching on the other hand (Czajkowski et al. 2014, Day et al. 2012, Caussade et al. 2005).

Overall, Chapter 2 as well as Chapter 4 found only some evidence for fatigue effects, at least in the range of the design dimensions investigated. Compared to research in transportation (Caussade et al. 2005), where a similar design approach was adopted, this dissertation rather provides evidence towards learning effects.

In line with research by Czajkowski et al. (2014) in an environmental context and Hess et al. (2012), who used datasets from different applications, decreases in the error variance particularly at the beginning of the sequence of choice tasks rather point to learning effects. Although it is not possible to disentangle institutional and preference learning, results from Chapter 2 and 4 suggest that both phenomena were present since the error variance was impacted over the entire sequence of choice sets. This raises the question whether responses to the first choice sets should be included or whether example choice sets should be used (see also Howard e l. 2017 and Carlson et al. 2012).

With respect to the number of alternatives, attributes and attribute levels, Chapter 2 (significant effects) and Chapter 4 (no significant effects) found, at the first glance, contradicting results. They, however, might be explained by differences in model specifications. While Chapter 2 employed a Heteroskedastic Logit model, Chapter 4 made use of a Mixed Logit framework. Czajkowski et al. (2014) note that different results in the literature might be explained by the model specifications. The results of this thesis, thus, suggest that there is no clear evidence that the number of alternatives, attributes and levels systematically influence choice consistency.

With respect to the level range, results of Chapter 2 and 4 are in line with previously published results (e.g. Caussade et al. 2005). Larger level ranges might have signaled to respondents a higher degree of uncertainty with respect to the magnitude of the environmental change to be valued. Preference uncertainty could also explain the result of Chapter 5, where choice consistency was positively affected when attribute thresholds were elicited prior to the DCE.

This thesis has further found evidence in line with the rational-adaptive model, i.e. impacts on the deterministic part of the utility function (DeShazo and Fermo 2004).

The strongest evidence in this regard is probably provided by Chapter 4 where welfare estimates calculated from designs which varied in term of complexity were observed to be significantly different. Similar findings were presented, among others, by Boxall et al. (2009). The fact that the status quo parameter had a particular large impact on these results emphasizes the need for guidelines on whether the status quo parameter should be considered or not when calculating non-marginal welfare measures (Adamowicz et al. 2011). A possible reason for detecting significantly different WTP estimates in Chapter 4 could have been the application of decision heuristics.

Chapter 3 and 4 showed that status quo choices increased as well as the attendance to attribute decreased, the more choice sets are presented. This result lends some evidence to the possibility that decreases in the error variance could be the result of adopting non-compensatory decision heuristics (Czajkowski et al. 2014). Therefore, the decreasing error variance observed in Chapter 2 and 4 might not only be explained by institutional learning, but could also be due to the fact that some respondents choose the status quo option more frequently towards the end of the DCE or to not attend to all attributes. However, due to identification problems effects on the deterministic part and the error structure of the utility function could not be accounted for simultaneously (see also Zhang and Adamowicz 2011 who experienced similar problems).

Findings from Chapter 4 further suggest that status quo effects result from the similarity between alternatives rather than the number of alternatives. Confirming observations made by Swait and Adamowicz (2001) and Zhang and Adamowicz (2011) higher levels of entropy led to more status quo choices. In line with observations made by Adamowicz et al. (2011), Zhang and Adamowicz (2011) and Rolfe and Bennett (2009), Chapter 4 found evidence for preference matching effects. With more alternatives it might have been easier for respondents to find a program that better matches their preferences resulting in less status quo choices. On the other hand, ANA was observed to increase with the number of alternatives (Chapter 3).

Intestinally, neither ANA nor status quo choices were impacted by the number of attributes. This finding may, however, be explained by the specific design approach used. Increases in the number of attributes have been achieved by disaggregating the attribute concerned with biodiversity. This approach was employed in order to be able to systematically account for differences in WTP estimates with respect to a common-metric attribute (Hensher 2004). As a consequence, respondents may have jointly evaluated the biodiversity attributes rather than considering each in isolation. In this context it has to be noted that Chapter 3 did not investigate ANA regarding the biodiversity attribute(s). Other studies such as Boxall et al (2009) provide evidence the number of attributes might influence preferences.

Chapter 4 also indicates that participants are less inclined to choose the current situation when the level range is less extreme. One might speculate that a reason for this is the use of attribute cut-offs (Swait 2001). When the level range became wider, some level values may have exceeded the maximum (minimum) acceptable level. As a consequence, respondents remained at the status quo to avoid cut-off violations.

Results from Chapter 5, indeed, indicate that attribute cut-offs play a role in making choices and that accounting for such cut-offs impacts model estimates quite substantially pointing again to non-compensatory decision making. However, respondents are in many cases willing to violate their cut-offs when having to make trade-offs. Similar observations were made within an environmental context (Bush et al. 2009) as well as in other DCE applications such as food choice (e.g. Ding et al. 2012) or transportation (e.g. Roman et al. 2017). These studies and Chapter 5, however, assume attribute cut-offs to be exogenous to the choice process. As pointed out by Ding et al. (2012), for example, cut-off endogeneity is potentially an important issue.

Chapter 5 further found that eliciting cut-offs prior to a DCE might impact respondent's preferences. An explanation for this result may be respondents' unfamiliarity with some of the attributes of renewable energy alternatives, which might have led respondents to learn about their preferences based on attribute constraints. Swait (2001) already pointed out that asking participants on their attribute constraints before starting the DCE might entail the risk that preferences are formed based on these cut-off rather than on past experience.

Lastly, results from Chapter 6 do not provide evidence for response anomalies resulting from respondents' perceptions of survey consequentiality. Unlike previous research (Vossler et al. 2012, Herriges et al. 2010, Nepal et al. 2009), the survey underlying Chapter 6 employed multinomial choice questions with four labelled alternatives while the literature has so far used choice sets with two alternatives. The only exception is Interis and Petrolia (2014) who found behavioral differences between a binary and a multinomial choice task when taking perceived consequentiality into account. This points to the need to further investigate the incentive structure of DCEs with more than two alternatives and a series of choice tasks (see also Section 7.3).

A consequentiality device, which was found to have no impacts in Chapter 6, has been observed to influence WTP estimates in other studies (Landry and List 2007, Bulte et al. 2005). One explanation for this could be that respondents did not recognize the information treatment in the survey. There is no measure to confirm or reject this. The amount of time participants spent on the treatment page, however, suggests that the additional information was read. It also has to be noted that all respondents were told on the first screen of the questionnaire that the objective of the survey was to find out what people in Germany think about the expansion of renewable energies, and that the results would be used to inform decision-makers. The consequentiality device was found to have some impact on consequentiality perception, but the impact of other factors such as respondents' interest in the topic, and, particularly, trust in institutions were observed to have much stronger effects on perceived consequentiality. This result emphasizes the limited ability of the researcher to promote consequentiality as perceived consequences of the survey might be driven by many factors (similar findings were presented by Vossler and Watson 2013).

7.3 Implications for Designing Discrete Choice Experiments

An important step in designing a DCE is to determine the design dimensions of the DCE. Increasing the complexity of the DCE, particularly the number of choice sets and alternatives, is advantageous from a statistical point of view since it increases the number of data points obtained from each respondent.

Since the study context and processing strategies adopted may vary a lot across applications and respondents, respectively, it will not be possible to determine a choice design which always minimizes the risk of biased estimate. Nevertheless, results from Chapter 2, Chapter 3 and Chapter 4 may be used to derive at least some recommendations for researchers and practitioners. First of all, it can be concluded from Chapter 2 that survey drop-outs due to variations in the design dimensions of the DCE are of minor concern. Although increases in choice task complexity increased the probability of pre-maturely abandon the survey, using more choice sets may still be advantageous in the light of the cost of statistical information, i.e. the cost per interview. As Louviere et al. (2013) argue, the negative effect of designs with a higher dimensionality on drop-out rates may be overcompensated by the additional information obtained. In fact, Chapter 2 illustrates this by calculating the number of data points obtained taking the example of designs with "low" and "high" complexity. Furthermore, respondents usually receive an incentive if they complete the questionnaire. Thus, participants might not quit the survey during the DCE since they have already paid the "sunk costs" of answering warm-up questions, etc.

With respect to impacts on model results the number of choice sets as well as the number of alternatives seems to be the most "critical" design dimensions. With the findings of this dissertation and other recent publications (Czajkowski et al. 2014, Hess et al. 2012) in mind, one might conclude that research and practitioners in environmental valuation could go beyond four to eight choice tasks, a number which is typically applied in DCEs for environmental valuation. A number of ten to 15 choice sets should not affect choice consistency negatively. However, Chapter 3 and Chapter 4 showed that more choice sets might lead to non-compensatory decision making. This should be kept in mind and accounted for when designing DCEs and analyzing choice data. Concerning the number of alternatives, DCEs for environmental valuation usually comprise three alternatives states of the world with one of these reflecting the status quo (Boxall et al. 2009). Chapter 4 suggests to use more than three or more alternatives to allow for a better matching of preferences. However, using more than two alternatives might raise issues with respect to the incentive compatibility of the valuation task, an issue that will be discussed in more detail below. Eventually, the number of alternatives to be presented will highly depend on the real-world context in which the DCE is employed (Johnston et al. 2017, Rolfe and Bennett 2009). For instance, in cases of labeled choices (e.g. renewable energy alternatives, recreation sites) it would be difficult to restrict toe choice set to a binary choice task.

Also, the level range should be determined with some degree of caution. One should be aware that larger distances between attribute levels may lead to less consistent and more status quo choices. Chapter 5 further suggests that research and practitioners should be aware that auxiliary survey questions prior to the DCE may have an influence on preferences. This could be particularly the case when respondents have to make statements on some of the attributes and their levels that are later used in the DCE.

From all these findings I, finally, recommend to use split sample approaches in DCE studies in order to be able to test for possible impacts of the survey design.

7.4 Future Research

Based on the discussion presented above, this section intends to point to future research. Chapter 2, 3, and 4 showed that choice task complexity may particularly impact the deterministic part of the utility function. Probably the most important and at the same time troubling practical discovery of Chapter 2, 3 and 4 was that the design of the DCE was observed to have significant influence on policy-relevant marginal as well as non-marginal welfare estimates. This empirical result, however, relates to only one dataset from a large-scale, nation-wide conducted online survey. Future studies could start by employing the DoD approach or other systematic measures of choice task complexity using a larger sample size. The total sample size of the dataset used to answer research question 1, 2 and 3 was 1,684 participants. This resulted in an average number of around 100 respondents per design. There is also a clear need to analyze impacts of task complexity in other study contexts. Rose et al. (2009), who applied the DoD approach in different countries in a transportation context, argue that the impact of design dimensions upon model estimates may be not necessarily transferable to other contexts. This may particularly apply to environmental valuation where the characteristics of goods and services differ a lot ranging from changes in biodiversity to valuations of the landscape view. One might also want to vary the number of attributes by a different approach than the one followed here. The problem, however, is not to confound effects of an increased number of attributes with effects due to information conveyed by attributes. Since the survey on land-use changes was conducted online, further paradata could be used to explain choices. So far, only survey drop-outs have been studied, but one could also use available data on the time respondents needed to choose an alternative. Campbell et al. (2017) found a clear link between response time and utility coefficients, error variance as well as processing strategies emphasizing the importance of considering response time when modeling stated choice data.

Another issue to consider is the inclusion of choice sets with only one hypothetical and a status quo alternative. It has long been known that DCE designs with two alternatives and a single choice set are incentive compatible (Carson and Groves 2007). Vossler et al. (2012) further showed that under certain conditions a sequence of binary choice tasks can be incentive compatible while the incentive compatibility of DCE with more than two alternatives has still to be demonstrated. Thus, including designs with two alternatives in a systematic analysis of choice task complexity as well as in a context where attribute cut-offs are elicited would enable further interesting insights.

Finally, future research endeavors should devote more attention to directly identifying and accommodating different decision strategies in discrete choice models. Research by Sandorf and Campbell (2018) or Daniel et al. (2018), who studied satisficing and elimination-by-aspects behavior using a two-stage model approach, moves in this direction. Whether status quo effects, ANA or the application of attribute cut-offs, this dissertation has shown that respondents in DCE make use of a variety of decision heuristics. This highlights the need to dedicate more research effort to information processing heterogeneity rather than preference and scale heterogeneity in order to bring DCEs and discrete choice models even closer to behavioral realism.

7.5 References

Adamowicz, W., Dupont, D., Krupnick, A., Zhang, J. 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods.

Boxall, P., Moon, A., Adamowicz, W.L. 2009. Complexity in Decision Making: The Effect of Task Structure and Context on Participant Response. Australian Journal of Agricultural and Resource Economics 53(4), 503–519.

Bulte, E., Gerking, S., List, J.A., de Zeeuw, A. 2005. The Effect of Varying the Causes of Environmental Problems on Stated WTP Values: Evidence from A Field Study. Journal of Environmental Economics and Management 49, 330–304.

Bush, G., Colombo, S., Hanley, N. 2009. Should all Choices Count? Using the Cut-Offs Approach to Edit Responses in a Choice Experiment. Journal of Environmental and Resource Economics, 44(3), 397–414.

Campbell, D., Mørkbak, M.R., Olsen, S.B. 2017. The link between response time and preference, variance and processing heterogeneity in stated choice experiments. Journal of Environmental Economics and Management 88, 18–34.

Carlsson, F., Mørkbak, M.R., Olsen, S.B. 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling, 5(2), 19–37.

Carson, R.T., Groves, T. 2007. Incentive and informational properties of preference questions. Journal of Environmental and Resource Economics 37(1), 181–210.

Caussade, S., Ortúzar, J., Rizzi, L.I., Hensher, D.A. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Ressearch Part B: Methodological 39(7), 621–640.

Czajkowssk, N., Giergiczny, M., Greene, W.H. 2014. Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. Land Economics 90, 324– 351.

Daniel, A.M., Persson, L., Sandorf, E.D. 2018. Accounting for elimination-by-aspects strategies and demand management in electricity contract choice. Energy Economics 73, 80-90.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63(1), 73–91.

DeShazo, J.R., Fermo, G. 2004. Implications of rationally-adaptive pre-choice behavior for the design and estimation of choice models. Working paper, School of Public Policy and Social Research, UCLA.

Ding, Y., Veeman, M.M., Adamowicz, W.L. 2012. The influence of attribute cut-offs and consumers' choices of a functional food. European Review of Agricultural Economics 39(5), 745–769.

Drechsler, M., Egerer, J., Lange, M., Masurowski, F., Meyerhoff, J., Oehlmann, M. 2017. Efficient and Equitable Spatial Allocation of Renewable Power Plants at the Country Scale. Nature Energy 2, 17124.

Haab, T.C., Interis, M.G., Petrolia, D.R. 2013. From Hopeless to Curious? Thoughts on Hausman's "Dubious to Hopeless" Critique of Contingent Valuation. Journal of Applied Economic Perspectives and Policy, 35(4), 593–612.

Hausman, J., 2012. Contingent valuation: from dubious to hopeless. Journal of Economic Perspectives 26(4), 43–56.

Hensher, D.A. 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. Journal of Transport Economics and Policy 38(3), 425–446.

Herriges, J., Kling, C., Liu, C.-C., Tobias, J. 2010. What are The Consequences of Consequentiality? Journal of Environmental Economics and Management 59, 67–81.

Hess, S., Hensher, D.A., Daly, A. 2012. Not bored yet – Revisiting respondent fatigue in stated choice experiments. Transportation Research Part A: Policy and Practice 46(3), 626-44.

Howard, B.E., Roe, E., Nisbet, C., Martin, J.C. 2017. Hypothetical Bias Mitigation Techniques in Choice Experiments: Do Cheap Talk and Honesty Priming Effects Fade with Repeated Choices? Journal of the Association of Environmental and Resource Economists 4(2) 543–573.

Johnston, R.J., Boyle, K.J., Adamowicz, W.L, Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C.A. 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists 4(2), 319–405.

Kling, C.L., Phaneuf, D.J., Zhao, J. 2012. From Exxon to BP: Has some number become better than no number. Journal of Economic Perspectives 26(4), 3–26.

Landry, C.E., List, J.A. 2007. Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field." American Journal of Agricultural Economics 89, 200–429.

Louviere, J.J., Carson R.T., Burgess, I., Street, D., Marley A.J. 2013. Sequential preference questions factors influencing completion rates and response time using online panel. Jounral of Choice Modell 8, 19–31.

Nepal, M., Berrens, V.P., Bohara, A.K. 2009. Assessing Perceived Consequentiality: Evidence from a ontingent Valuation Survey on Global Climate Change. International Journal of Ecological Economics and tatistics 14, 14–29.

Rolfe, J., Bennett, J. 2009. The Impact of Offering Two Versus Three Alternatives in Choice Modelling Experiments. Ecological Economics 68(4), 1140–1148.

Román, C., Arencibia, A. I., Feo-Valero, M. 2017. A latent class model with attribute cut-offs to analyze modal choice for freight transport. Transportation Research Part A: Policy and Practice 102, 212-227.

Rose, J., Hensher, D.A. Caussade, S., de Dios Ortúzar, J., Jou, R.-C. 2009. Identifying Differences in Willingness to Pay due to Dimensionality in Stated Choice Experiments: A Cross Country Analysis. Journal of Transport Geography, 17(1):21-29.

Sandorf, E.D. Campbell, D. 2018. Accommodating satisficing behavior in stated choice experiments. European Review of Agricultural Economics, forthcoming.

Swait, J. 2001. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodigical, 35(10), 903–928.

Swait, J., Adamowicz, W.L. 2001. Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. Organizational Behavior and Human Decision Processes 86(2), 141–167.

Train, K. 2009. Discrete choice methods with simulation. Cambridge University Press.

Vossler, C.A., Doyon, M., Rondeau, D. 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics 4, 145–171.

Vossler, C.A., Watson, S.B. 2013. Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field. Journal of Economic Behavior & Organization 86, 137–147.

Zhang, J., Adamowicz, W.L. 2011. Unraveling the Choice Format Effect: A Context-Dependent Random Utility Model. Land Economics 87(4), 730–743.

Statement of Contribution

Chapter 2, 3, 4 and 6 of this dissertation are the result of collaborations between the author, Jürgen Meyerhoff, Priska Weller and Petr Mariel. The author has made significant contributions to the content of all chapters, from the development of the research question, to the empirical analysis and the writing. Chapter 1, 5, and 7 were outlined and written solely by the author with Jürgen Meyerhoff providing helpful comments to Chapter 1, 5 and 7, Volkmar Hartje to Chapter 1 and 7 as well as Patrick Lloyd-Smith and Vic Adamowicz to Chapter 5.

The survey which founds the basis of Chapter 2, 3 and 4 was jointly designed and tested by the author, Priska Weller and Jürgen Meyerhoff. In close collaboration with Jürgen Meyerhoff and Priska Weller the author developed the research questions to be addressed with this survey.

The author, Jürgen Meyerhoff and Geesche M. Dobers jointly designed and tested the survey underlying Chapter 5 and 6. In close cooperation with Jürgen Meyerhoff the author developed the research questions which are addressed in this survey. The contributions of the author to Chapter 2, 3, 4 and 6 are detailed in the following:

Chapter 2: Meyerhoff, J., Oehlmann, M., Weller, P. 2015. The Influence of Design Dimensions on Stated Choice in an Environmental Context. Journal of Environmental and Resource Economics, 61(3), 385–407.

Jürgen Meyerhoff and the author jointly wrote the paper and conducted the data analysis and model estimation. The writing was supported by Priska Weller.

Chapter 3: Weller, P., Oehlmann, M., Mariel, P., Meyerhoff, J. 2014. Stated and inferred attribute non-attendance in a design of designs approach. Journal of Choice Modelling, 11, 4–15.

Priska Weller wrote the article with substantial contributions from the author. The author also contributed to the data analysis. The model estimation was realized by Priska Weller, Jürgen Meyerhoff and Petr Mariel.

Chapter 4: Oehlmann, M., Meyerhoff, J., Mariel, P., Weller, P. 2017. Uncovering context-induced status quo effects in choice experiments. Journal of Environmental Economics and Management, 81, 59–73.

The article was developed in close cooperation with Jürgen Meyerhoff who contributed through extensive discussion in all stages of the research. The author coordinated and wrote the article and conducted the data analysis. The modeling and estimation process was further supported by Petr Mariel. Priska Weller supported the writing with valuable comments.

Chapter 6: Oehlmann, M., Meyerhoff, J. 2016. Stated preferences towards renewable energy alternatives in Germany – do the consequentiality of the survey and trust in institutions matter? Journal of Environmental Economics and Policy, 6(1), 1–16.

The author coordinated and wrote the article with input from Jürgen Meyerhoff. In collaboration with Jürgen Meyerhoff, the author conducted the data analysis and model estimation. Geesche M. Dobers supported the article with helpful comments.

Overview of Publications

Chapter	Published as	Version
2	Meyerhoff, J., Oehlmann, M., Weller, P. 2015. The Influence of Design Dimensions on Stated Choice in an Environmental Context. Journal of Environmental and Resource Economics, 61(3), 385–407. https://doi.org/10.1007/s10640-014-9797-5	Accepted/Last submitted
3	Weller, P., Oehlmann, M., Mariel, P., Meyerhoff, J. 2014. Stated and inferred attribute non-attendance in a design of designs approach. Journal of Choice Modelling, 11, 4–15. https://doi.org/10.1016/j.jocm.2014.04.002	Accepted/Last submitted
4	Oehlmann, M., Meyerhoff, J., Mariel, P., Weller, P. 2017. Uncovering context-induced status quo effects in choice experiments. Journal of Environmental Economics and Management, 81, 59–73. http://dx.doi.org/10.1016/j.jeem.2016.09.002	Accepted/Last submitted
5	Malte Oehlmann: Do Attribute Cut-offs Make a Difference? The Effects of Eliciting and Incorporating Cut-off Values in Choice Models. Working paper presented at the 5th International Choice Modelling conference 2017 and the 6th World Congress of Environmental and Resource Economists	In Preparation
6	Oehlmann, M., Meyerhoff, J. 2016. Stated preferences towards renewable energy alternatives in Germany – do the consequentiality of the survey and trust in institutions matter? Journal of Environmental Economics and Policy, 6(1), 1–16. http://dx.doi.org/10.1080/21606544.2016.1139468	Accepted/Last submitted

Acknowledgement

First of all, I would like to thank Volkmar Hartje for offering me the opportunity to work and conduct my PhD at the Chair of Environmental and Land Economics at TU Berlin. I also want to thank him for his patience and for supporting my dissertation with guidance, freedom and trust.

I want to express my gratitude to Katrin Rehdanz who volunteered to be the second advisor of this thesis. Special thanks go to Jürgen Meyerhoff for many, many fruitful and extensive discussions, guidance throughout my whole dissertation and the day-to-day supervision he provided. Thanks Jürgen for introducing me to the choice modeling world!

I am also grateful to Vic Adamowicz for offering me the opportunity for an inspiring research stay at the University of Alberta. Thanks also to Patrick Lloyd-Smith for many interesting discussions and helpful comments during my time in Canada.

Special thanks also go to:

- Priska Weller and Geesche M. Dobers for a great collaboration when preparing analyzing the surveys
- Henry Wüstemann for, amongst others, several off-topic conversations when sharing the same office
- Julian Sagebiel for many helpful comments when finishing my PhD
- Katharina Klaas and Miró Mandelkow for assisting me with the layout of the articles and the final document

Lass but not least, I would like to thank all of my co-authors and colleagues: Thank you for a wonderful working atmosphere, it has been a pleasure to be and to work with all of you!