New research program devoted to enhancing the validity and robustness of stated preference valuation methods in the face of selected behavioral phenomena and biases

1. Research objectives

1.1. Introduction

The proposed project will address state-of-the-art methods for **modeling consumers' preferences** and the **valuation** of non-market goods. Preference modeling underlies modern microeconomics, and it is difficult to overemphasize its importance. Preference modeling is critical for determining socially optimal levels for the provision of public goods, designing new private goods with the most desired characteristics and predicting consumer behavior. Without knowing consumer preferences (and not being able to observe their market decisions), valuation and subsequently, the efficient allocation of resources, would not be possible (e.g., Freeman et al. 2014, Champ et al. 2017).

The theory of economics has developed several methods that estimate the value of non-market goods. The first group of these methods is based on consumers' **Revealed Preferences** (RP) – the data on actual choices individuals made in market situations.[1] The second group of the valuation methods is based on respondents' **Stated Preferences** (SP) – data are obtained through adequately designed surveys that include hypothetical choice situations (Hanley and Czajkowski forthcoming). For example, suppose we are interested in how much people would pay for developing some new road connection. This could be measured using a **Discrete Choice Experiment** (DCE; Carson and Czajkowski 2014)[2], where the good is described as a bundle of individually varied and separately valued attributes.[3] The possible development scenarios can be described using a set of attributes of the new connection – the number of lanes, types of crossings, expected travel time, and cost to their household (incurred in the form of taxes or toll fees). The choices observed in these hypothetical situations make it possible (with the use of the appropriate statistical methods) to estimate the utility function parameters that are related to the specific attributes of a good and hence formally model consumer preferences. Therefore, it is possible to evaluate changes in consumer welfare in the case of implementing of a particular scenario (e.g., providing a new public good) and predict consumers' behavior that is related to new goods or alternatives. In addition, identifying the marginal rates of substitution between particular characteristics of a good (including the pecuniary attribute, e.g., the cost for respondents' household) makes it possible to identify people's WTP for non-market goods and their characteristics.

The RP and SP methods have been implemented in a vast number of papers in many fields of applied microeconomics, including marketing, transport, health and environmental economics. This can be illustrated by over 7,500 papers and studies referenced in Carson (2012a) utilizing the method (with the largest group focused on environmental valuation) and with exponentially growing number of currently over 185,000 citations of papers applying and considering these methods, as evidenced in Hanley and Czajkowski (forthcoming). The discussion of RP and SP methods is now standard in almost all textbooks on environmental economics. Estimates derived from stated preference data constitute almost 60% of the estimates in Environment Canada's very large Environmental Resource Inventory (EVRI) database maintained in

[1] One example of these techniques is the **Travel Cost Method** (TCM; Parsons 2017), in which the frequency and cost of visits to a site (e.g., a national park) are used to estimate the demand curve and consumer surplus from being able to access it. Another prominent example is the **Hedonic Price Method** (HPM; Taylor 2017), in which differences in prices of goods are associated with differences in their selected characteristics (e.g., environmental amenities of a house, accident risk and wage rate, organic or fair trade labels of agricultural products).

[2] We note that the literature is divided over nomenclature. Whilst, historically, discrete choice experiments (DCE) were proposed as an alternative approach to contingent valuation (CV), argue that they are a form of elicitation format used in CV studies. As a result, not all CVs are DCEs (e.g., when they use non-discrete choice formats) and not all DCEs are CVs. While DCE may be a part of CV, a single choice format studies (yes or no to a new policy at a cost, using payment card, etc.) are usually referred to as CVs, while studies that use many choices and/or or a choice from many alternatives for a single respondent are usually referred to as DCEs.

[3] In alignment with the Lancasterian perspective of utility (Lancaster 1966), every good is described in terms of a collection of its characteristics (attributes). The selected combinations of levels of these attributes include the alternatives that are presented to respondents, who are asked to choose the alternative that they consider the best (the most preferred).

conjunction with environmental protection agencies in several other countries.[4] SP studies are used for assessing very large policies, such as the U.S. Clean Water Act (US Environmental Protection Agency 1994), and for a wide range of policy decisions (see, e.g., Griffiths et al. 2012, Atkinson et al. 2018 for a recent review).

1.2. Statement of the research problem

Almost from the start of their use in environmental economics, SP valuation methods occasionally turned up behavior which was thought to be potentially at odds with standard neoclassical economic theory describing consumer choice and welfare measurement (Carson and Hanemann 2005). Therefore, the issue of **hypothetical market bias**[5] remained a key criticism of SP methods until a new stream of improvements of the method emerged, drawing insights from mechanism design theory. The aim was making the statement of respondents' true preferences their best available strategy: that is, to make survey questions **incentive compatible**.[6] In SP surveys this can be achieved by refining the construction of hypothetical scenarios, making the surveys appear consequential, using coercive payment mechanisms and following several other conditions (e.g., Carson and Groves 2007, Vossler et al. 2012, Czajkowski et al. 2017).[7] In this setting, the observed anomalies can no longer be attributed to hypothetical bias.

Some of such behavioral phenomena, were later shown to be quite robust across a range of non-market and market situations, and to exist well beyond the ambit of SP applications.[8] In some cases, however, their existence can question the validity of SP methods, because they can make the resulting value estimates design-specific. **The aim of this project is to systematically investigate such biases and propose new ways to deal with them, thus making SP methods more robust and enhancing their validity.**

One of the behavioral phenomena investigated in this project, that is particularly worrying for the validity and robustness of SP methods, is the **anchoring effect** (Ariely et al. 2003). In simple terms, anchoring with respect to the selection of the bid vector to be used in SP studies can make respondents focus on the bids that they see, when establishing their WTP, and hence by choosing a lower or higher set of bids a researcher **could arbitrarily influence the resulting WTP**.

Early studies that noticed this effect were concerned with (optimal) bid selection in simple CV studies (e.g., Cooper and Loomis 1992, Cooper 1993, Boyle et al. 1998, Bateman et al. 2001). For example, Veronesi et al. (2011) investigated the issue of a starting point bias in double bounded discrete choice questions, Day and Pinto (2010) looked at general ordering effects of choice tasks, while Meyerhoff and Glenk (2015) and Ladenburg and Olsen (2008) studied the effects of an instructional choice set on respondents' choices and implied preferences and found that starting point matters and its effect can be respondent-specific (depending on, e.g., respondents' gender and other socio-demographic characteristics).

Other studies investigated sensitivity to changing attribute and cost levels more directly. **The results of some of these studies are supportive for the use of SP methods**, as they find they are unaffected by the selection of the bid levels vector. For example, Hanley et al. (2005) found a rational response to changing the design

[4] https://www.evri.ca/.

[5] Simply put, this means people over- or under-state their WTP values in an SP exercise in a systematic fashion due to the fact that no actual payment is made or received in exchange for an actual change in quantity or quality of a good.

[6] Incentive compatibility of SP methods has been another area of interest and research of the leader of this project.

[7] Hypothetical bias has been thoroughly investigated since, particularly because the empirical evidence is often contradictory. Some studies report significant differences between stated and true preferences, whereas others find no significant difference. Recently, Zawojska and Czajkowski (2017) have critically re-evaluated this evidence. By reviewing the four main types of validity tests – content, construct, convergent, and criterion validity – they argue that comparing SP-based estimates with corresponding criterion measures is the most adequate approach to verify how well SP-based estimates reflect true preferences. By classifying the empirical evidence with respect to whether it (1) deals with private or public goods, (2) uses a coercive or voluntary payment mechanism, (3) can be perceived by respondents as consequential, and (4) uses a single binary choice format, they identified studies that provide meaningful results in terms of providing conditions in which rational respondents can be expected to answer in line with their true preferences. The results of such studies consistently point to the validity of stated preferences under such conditions. When the available evidence is limited only to studies that satisfy the requirements listed above, the evidence becomes univocal – hypothetical bias can be avoided. This conclusion is very encouraging for SP methods, although it obviously comes with many requirements for the design and administration of future SP studies.

[8] For example, systematic differences between values attached to gains compared to losses (WTP-WTA disparity; Knetsch and Sinden 1984, Coursey et al. 1987, Boyce et al. 1992).

price vector, in that higher prices resulted in a lower rate of acceptance of costly alternatives to the status quo. Once differences in variance between samples had been allowed for, changing the price vector had no significant effect on estimates of preferences (based on likelihood ratio tests of the underlying probabilistic choice models).[9] Aravena et al. (2014) compared the effects of framing the changes as price increase or decrease relative to current level, or using no price in the experiment. Their results support the neoclassical theory as they found that the means of the conditional estimates of the marginal values of attributes were unaffected by the direction of the price change and the exclusion of the price attribute.[10]

Another group of studies reported somewhat inconclusive results. For example, Carlsson and Martinsson (2008) found a significant effect for the additive scaling of the cost vector, although they did not observe a significant effect of the first choice task presented to respondents.[11] Ryan and Wordsworth (2000) assessed the sensitivity of WTP-estimates to changes in multiple attribute levels. Two split-samples were administered with the levels varying for three out of six attributes, including the cost attribute. Their results indicated a significant impact of varying attribute levels on mean WTP-estimates. However, because the ranges in multiple attribute levels were changed simultaneously, the found effects cannot be contributed solely to a change in cost levels. Similarly, Svenningsen and Jacobsen (2018) compare the effects of using a low or high bid range, but because using different bid ranges was confounded with using different payment mechanisms (hypothetical tax and donation vs. real donation, respectively) their results cannot be univocally interpreted as the anchoring effect. Glenk et al. (forthcoming) observed that the acceptance rates of alternatives are only marginally decreasing with increasing magnitude of cost vectors. This was found to be the case especially for larger cost vector levels despite the fact that absolute differences in cost were increasingly large for these levels. Given this pattern in bid acceptance, a clear increase in WTP would be expected as the magnitude of cost vectors increases. However, differences in marginal WTP between cost vector sub-samples were relatively modest and only significant for few attributes and comparisons across cost vector treatments.

Finally, some studies observe clear effects of anchoring in DCE studies. Hensher (2004, 2006) reports that using wider or narrower attribute level range significantly influenced WTP, at least for some designs. Mørkbak et al. (2010) investigated the choke price bias in DCE by including an increasingly large amount in the bid vector and found that it significantly influences the observed WTP. Kragt (2013) observed that the proportion of respondents who chose the no-cost base option over costly environmental management alternatives between the 'standard cost' and 'high cost' questionnaire versions was not significantly different. The distribution of responses was relatively insensitive to the absolute price vector, indicating the presence of an anchoring effect. Su et al. (2017) conclude that the models with low price levels and high (doubled) price levels are different. Similarly, Burrows et al. (2017) observe statistically different results when using higher (quadrupled) bid levels.

Overall, the evidence for anchoring (and similarly, for some other behavioral phenomena) in SP studies appears mixed. **Not much is generally known about the nature and circumstances in which behavioral effects, such as anchoring, manifest themselves, potentially jeopardizing the validity of SP methods.** With the rare exceptions of Sugden et al. (2013), who investigate what anchors are the strongest for the observed WTP and WTA for private goods and Caputo et al. (2018), who show how the level and variability of price perceptions "outside" the task at hand might affect choice behavior for private goods, **many other questions remain unanswered,** for example: Why is anchoring observed in some SP studies and not observed in others? Is this related to the construction of the study (including its incentive compatibility properties), respondents' experience with the good (c.f. Ahtiainen et al. 2014, LaRiviere et al. 2014) or its nature (public, quasi-public, private)? Luisetti et al. (2011) hypothesize, that respondents may perceive attribute levels in a relative rather than absolute sense; they use a split sample treatment of relative or absolute distance to demonstrate its effect. Can similar mechanism take place for the cost and other attributes typically used in SP studies? If respondents do anchor – what do they anchor to: the first bid they see, the average (mean) levels of bids, the maximum bid used in the experiment? What cues does the variation of the bid levels provide to individuals and how does it influence observed WTP? To what extent does low quality of some responses (e.g., speeding through surveys,

[9] Although moving to the lower price design resulted in a reduction in mean willingness-to-pay for discrete improvements in all three river quality attributes, these changes were not statistically significant.

[10] However, they did observe a larger spread of values when the choice experiment implied a tariff decrease, than an increase (c.f. Bartczak et al. 2017).

[11] Importantly, their experiment did not include an opt-out alternative.

survey modes promoting low attention) and misspecification of the model (such as not accounting for respondents' preference heterogeneity, or the discrete-continuous nature of some decisions) contribute to the observed effects? Can the observed anomalies be avoided by combining SP and RP data sources? Can changes in survey construction help in developing home-grown, rather than anchored responses? **These are all important research questions and research gaps that our project will address**.

2. Work plan

The research program designed for this project consists of exploration of several issues related to behavioral anomalies observed in SP studies. By studying these issues we will be able to investigate possible causes for the behavioral anomalies observed in SP studies and propose ways to avoid them or alleviate their consequences.

2.1. Private and public good nature

We intend to investigate if (while controlling for **experience and information** with the good, typically larger for private goods) the **behavioral phenomena** jeopardizing the validity of SP methods are **more pronounced for public than for private goods (**Brown et al. 2008**)**. This is not a trivial question, as designing incentive compatible SP choice scenarios are more difficult for private goods, while the most of the existing evidence comes from this setting (Vossler and Holladay 2018). This simple work package will potentially yield a valuable insight into the extent to which the observed anomalies are driven by the experience, information, preference elicitation format and the characteristics of the hypothetical choice scenario (typically different for private and public goods).

2.2. Quality of survey responses

One stream of research dealing with the economics profession's worries about behavioral biases in SP studies focused on **improving survey design methods**, and **developing ex-ante / ex-post methods** to reduce behavior that appears not in line with the economic theory (Carlsson et al. 2016). Such methods include cheap-talk scripts, oath-scripts, time-to-think, certainty-calibration, budget reminders, opt-out reminders and other. It quickly became obvious that poor survey design and poor administration could easily induce all sorts of anomalous behaviors. On the other hand, SP studies that invested considerable time and effort into understanding what people currently understand about a good, in presenting a credible choice scenario with a well-defined good and a coercive payment mechanism, and where survey design enhances belief in outcome and payment consequentiality, generally appear to produce results that are well-behaved.

In this work package we intend to follow up on this stream of research, and investigate if behavioral problems, such as the **"fat tails" in WTP distribution** (Parsons and Myers 2017), are related to the **survey administration mode** (Menegaki et al. 2016), "**speeding**" and generally **time** spent considering survey questions (Börger 2016), **attribute non-attendance** (Caputo et al. 2018, Sandorf et al. forthcoming) and other quality-related behaviors of survey respondents. In addition, a promising avenue for this research is utilizing "**trap-questions**" (Jones et al. 2015) and other forms of controlling for "**yeah-saying**" (Blamey et al. 1999, Holmes and Kramer 2002), the respondents' **attention** and **cognitive burden** (Swait and Adamowicz 2001).

2.3. Accounting for preference heterogeneity

In this and the three following work packages we intend to investigate the extent to which the cases behavioral anomalies reported in SP studies may be **a result of econometric model misspecification**. The question of how to correctly specify a choice model in order to obtain a good description of individuals' preferences has attracted a significant attention in the literature. The state of the practice of applied discrete choice analysis is based on multinomial logit, which was further extended to mixed logit (MXL; Revelt and Train 1998). This specification allows to account for unobserved preference heterogeneity in the form of random parameters. Although MXL is widely popular, it is not free of several shortcomings, such as the necessity to arbitrarily choose parametric distributions for preference parameters, and the necessity to resort to simulation methods to estimate the model. In the case **preference heterogeneity is not accounted for**, the selected **parametric distribution is far from the truth**, or is **insufficiently flexible** – the results can incorporate a wide array of anomalous artefacts. For example, Figuerola et al. (2014) provide a Monte Carlo simulation evaluating the extent of error in welfare measures resulting from incorrect assumptions regarding preference heterogeneity,

and Knittel and Metaxoglou (2014) review common errors, when using maximum simulated likelihood method to estimate MXL models. We believe that in some cases, these problems **can manifest themselves as apparent behavioral anomalies**, while they are in fact a result of the model misspecification. Using several available datasets we intend to systematically examine this issue. This will allow us to give recommendation regarding a use of this approach in applied welfare analysis using DCE.

2.4. Attribute Non-Attendance

The basic specification of a typical conditional multinomial choice model assumes that individuals are rational and pay attention to all the information presented to them in a DCE. Nevertheless, there is some evidence that when making choices some **individuals may use simplifying heuristics** to make a task at hand easier (Scarpa et al. 2013). One of such heuristics is **Attribute Non-Attendance (ANA)**, in which some individuals may ignore one or more attributes. There are two main ways in which ANA in usually implemented in the choice model. The first, usually called **inferred ANA**, employs statistical methods in which researcher is trying to discover the share of individuals who have a parameter equal to zero for a given attribute. This involves using a latent class approach in the model (Hensher and Greene 2010), which is directly connected with preference heterogeneity issue, as incorporating ANA adds to an assumed distribution a probability mass at zero. Because adding ANA leads to more flexible preference distribution, the identified effect may, in our opinion, be indicating incorrect assumptions regarding preference heterogeneity distribution. The second approach is the so called **stated ANA** (Hole et al. 2013), in which individuals are directly asked in the survey whether they ignored a given attribute when making choices. Although this approach is simpler from the computational point of view, it results in attribute levels that are not in line with the results of the inferred ANA (e.g., Hess and Hensher 2010). We believe this could be a result of using a model that does not account for endogeneity of the stated responses and the actual choices, leading to misleading conclusions. In addition, we intend analyze the effect of DCE design (such as number the of alternatives and the values of bids used in the experiment) on ANA.[12] In particular, ANA of the monetary attribute (e.g., cost) is a significant issue, as no welfare measures can be calculated for the individuals who ignore cost (Campbell et al. 2012). As a result, the relationship between bid vector and ANA is crucial for the validity of DCE. Overall, we intend to investigate if **not accounting for ANA, or accounting for it in an insufficient or incorrect way, may be a driver of various apparent behavioral phenomena reported by SP studies**.

2.5. Non-constant marginal utility of money

Nearly all econometric treatments of SP data assume **constant marginal utility of money** (or cost). This assumption is not supported by the empirical evidence. For example, Gunn (2001), Fox et al. (2009) and Mabit et al. (2013) observe the so called **"cost damping" effect**, which is characterized by diminishing marginal utility of cost (Daly 2010). Not accounting for this effect may lead to biased welfare change estimates (Budziński 2015) or inaccurate forecasts (Daly 2010), but also likely lead to anomalous results which can incorrectly be interpreted as other behavioral phenomena (e.g., anchoring, fat-tails problem, insufficient sensitivity to scope). We therefore intend to investigate the cost damping effect in the context of DCE. Specifically, we will analyze the **connection between the magnitude of the cost damping and the bid vector** that is used in the design of DCE. Some studies, connect this issue with "preference formation" (Stathopoulos and Hess 2012) – different anchoring points may lead to a different reference points, which will change non-linear sensitivity to the cost. The other issue which can be potentially connected with non-linear sensitivity to cost is the **confounding**, which arises when cost attribute takes value of zero only for status quo alternative, and, therefore, the marginal effect of this level cannot be distinguished from 'status quo' effect (Hess and Beharry-Borg 2012). We will investigate, whether a design in which these two effects can be separated will allow to observe well-behaved cost sensitivity.

2.6. Using discrete-continuous choice, rather than discrete choice models

The simple discrete choice models that are commonly used may not be appropriate for some consumer decisions. Many of these decisions are in fact **joint discrete-continuous choice processes**, in which a consumer chooses not only which goods to consume (possibly more than one), but also the quantity of each of

[12] The former increases the complexity of the choice task (Hensher et al. 2005a), and therefore we expect higher shares of individuals employing ANA heuristic. The latter may have different effects, depending on the bids chosen. For example if individuals do not consider a given bid value as credible they may choose to ignore it.

them. Examples of such situations can be found in the transportation literature (e.g., Bhat and Sen 2006, Ahn et al. 2008) or the recreation literature (von Haefen and Phaneuf 2003, von Haefen et al. 2004). To date this kind of models received very limited attention in literature, relative to the simple discrete choice models. In this project we aim to compare the results obtained from the joint discrete-continuous models with other approaches used in the literature, such as count data models (Czajkowski et al. 2015) and regular discrete choice models (Termansen et al. 2013).[13] We will focus on differences in estimated welfare measures between different approaches on revealed preference data, and particularly, on whether using the more appropriate discrete-continuous specification allows for accounting for **econometric model misspecification** that can manifest itself as various behavioral anomalies. All the software codes for our new econometric models will be made available as tools for future applications by other researchers, through the https://github.com/czaj/DCE GitHub platform under Creative Commons BY 4.0 license.

2.7. Experimental investigation of anchoring

In this work package we intend to **empirically explore** the phenomenon of anchoring in SP studies, by **experimentally varying the attribute levels in treatments** administered to separate groups of respondents. By varying the vectors of bids and other (non-monetary) attributes, we intend to investigate if the problem of anchoring refers to the mean, variance, range, minimum, maximum or possibly the first attribute level seen by respondents. This can allow for better understanding of how anchoring works, and in the case anchoring is unavoidable – selecting the attribute levels people anchor to in a way that closely mimics the levels considered by a new provision policy of the studied good.

In addition, in this work package we will also **test the methods to alleviate anchoring developed by our project**, such as using prior prompts to elicit respondents home-grown, rather than anchored preferences. One of such mechanisms could be asking an open-ended question, before moving to indirect (discrete choice or referendum type) elicitation formats in the same survey.

2.8. Convergent validity of SP and RP valuation methods

Although there is a growing interest in using SP methods for valuation of non-market goods there are still potential benefits of using RP datasets. In particular, combining the two sources of data may lead to better understanding of individual choice processes and enrich explanatory power of the model (Hensher and Bradley 1993). The strengths of RP data are usually the weaknesses of SP data and vice versa, therefore they can be used as complementary methods rather than substitutes (Whitehead et al. 2008). **Combining different sources of data** demands specific model formulation, in order to account for potential differences in scale and state dependence (Bhat and Castelar 2002, Hensher et al. 2008), which can occur due to different nature of data. In this work package, we aim to **use RP data** in order to **gain a better understanding of selected behavioral issues reported in some SP studies**. For example, we intend to test if by changing the different dimensions of DCE design, such as a bid vector or the number of alternatives, we can obtain results that are closer to the preferences inferred from RP data (and vice versa). Although the idea of comparing RP and SP data for validation of one or the other is not new (e.g., Azevedo et al. 2003) we propose a more thorough investigation, which will allow us to conclude which characteristics of a DCE study should be calibrated in order to mitigate potential biases. As a result, we will be able to gain new insights about validity of SP valuation methods, and their robustness to various behavioral phenomena that can find their way to hypothetical choice data.

2.9. Choice set formation

In a DCE, it is usually assumed that individuals behave rationally, taking into account information regarding all alternatives and attributes presented to them in a choice task. However, research in psychology and behavioral economics suggest that this may not always be the case (e.g., Swait 2001). Similarly as in the case of ANA, **individuals may not consider all available alternatives** in order to simplify decision process and reduce cognitive burden. This effect, usually referred to as "**choice set formation**", was **found in both, RP studies** (e.g., Banzhaf and Smith 2007, von Haefen 2008) **and SP studies** (Swait 2001, Thiene et al. 2017). Not taking this effect into account will lead to biased parameter estimates and welfare measures that are

[13] This was partially done by von Haefen and Phaneuf (2003), who compared multiple discrete-continuous models with count data models, but we will expand this comparison to discrete choice models, and use the novel, much more flexible formulation of the joint discrete-continuous models (Bhat 2005).

calculated on their basis (Li et al. 2015). Although this phenomenon is already established in the literature, there was little effort dedicated to better understanding this issue in DCE. As this issue arises in RP studies as well, we do not expect that such behavior is completely absent in SP. Nevertheless, it is likely that some decisions regarding DCE design may influence the extent of such behavior. For example, the basic model proposed by Swait (2001) assumes that certain alternatives do not enter an individual's choice set if attribute levels do not exceed a certain cut-off points. As levels of attributes in DCE are chosen by a researcher, various behavioral anomalies (such as framing and anchoring effects) could be a possible drivers of choice set formation behavior, for example, because they may determine cut-off points individuals use. In this work package we intend to analyze this issue and investigate the extent, to which it can contribute to the observed anomalous results of SP studies.

2.10. Synthesis of the results, conclusions and dissemination

The results of this research program will make it possible to verify our research hypotheses and confirm the feasibility and efficiency of the proposed methodological solutions. Our work will make it possible to prepare **at least 6 highly-visible and widely-cited scientific publications in the best journals of the field**, such as *Journal of Environmental Economics and Management*, *Journal of Choice Modelling*, *Journal of Consumer Research*, *Environmental and Resource Economics*, *Ecological Economics*, *Land Economics*, *American Journal of Agricultural Economics*, *Health Economics*, *Transportation Research* or *Marketing Science*. In addition, we intend to disseminate our results in other ways, such as **presentations at the most prestigious international conferences** (e.g., *Conference of the European Economic Association*, *Econometric Society*, *World Congress of Environmental and Resource Economics*, *International Choice Modeling Conference*, *Transportation Research Board Annual Meeting*, *World Congress of the International Health Economics Association*). The data and research collected within this project will make it possible to continue the analysis in the future and prepare additional publications. Finally, we believe that making the software codes developed within the project available to other researchers and practitioners as easy to use tools will substantially increase the scientific impact of our project and the visibility of our results, leading to a wider dissemination of the results (and more citations of our papers).

***

Overall, this project constitutes a research program consisting of nine components that are potentially contributing to understanding of the behavioral anomalies observed in SP studies, and can lead to addressing the problems these anomalies cause for the interpretation of SP results. These components constitute research agenda for me and the new members of my research team for the following five years. Pursuing each of these research ideas can lead to clear value added with respect to the current state-of-the-art, providing new evidence and developing innovative methodological approaches to modeling consumers' preferences and the valuation of non-market goods. I believe this research program has a chance to be ground-breaking, by freeing the state-of-the-art SP studies of the most burning problems associated with the observed behavioral phenomena and biases that could lead to invalidating modelling consumer preferences and valuation.

3. Methods of research

3.1. General approach and data sources

The project deals with using SP valuation methods – designing and constructing SP studies, data treatment and analysis methods, interpretation of the results. The project will be conducted in accordance with the work plan specified in the proposal. Our research is **work-intensive** (research ideas, methods and verification and econometric methods to be developed and used in this project were described in the full proposal), but on the other hand, most of the research components can be completed using **existing data sets** obtained in our earlier studies as well as studies conducted by other researchers, who make the data for their published studies available. We therefore only plan **two empirical studies that require collecting new primary data** – field DCE surveys – devoted to experimental investigation of anchoring (varying the attribute levels in treatments and testing the methods to alleviate anchoring developed by our project). The surveys will be prepared in accordance with the state-of-the-art recommendations for SP studies (e.g., Johnston et al. 2017). They will include carefully designed treatments, which will allow to verify our research hypotheses and provide insights

on the research questions of the project. The samples used in the research will be representative for target populations (affected by the policies / public goods in question). The surveys will be administered by professional public opinion research companies (subcontracted in accordance with the Polish public procurement law). The context of the empirical studies will be selected in such a way that the empirical results, apart from enabling us to verify research hypotheses formulated in the project, will prove interesting for those responsible for shaping policies in selected aspects of social life, because they will make it possible to determine the rational (i.e. maximizing the social welfare) level of providing public goods and their characteristics which consumers desire the most (for example, the value of travel time in the context of transportation). The applied aspect of these studies can be used as a basis for stand-alone publications targeted at field journals (e.g., in transport economics, environmental and resource economics, health economics, culture economics).

3.2. International cooperation

The project will be conducted in cooperation with other scientists from abroad, who will be invited to participate in selected streams of investigation and co-author joint publications. In the past, most of my research was conducted with such international collaborators, who co-authored the resulting papers. I believe such cooperation allows one to be more efficient, specialize, learn new skills and generally aim higher. In addition, I believe such international cooperation will be particularly valuable for other participants of the project (junior researchers, PhD students), because it is an efficient way to promote their scientific development and build their personal relationships with top scientists in the field. I have no doubt participation in the research tasks specified in the work plan of this proposal will be interesting for top researchers in respective areas, as they constitute the hottest topics in the current state of research in these areas.

3.3. Econometric approaches: random parameters multinomial choice (MXL) models

Theoretical foundations for quantitative modelling of consumers' utility functions are provided by the **random utility theory** . The random parameters model is based on the following logic: respondents *i*'s utility associated with choosing alternative *j* out of the *J* available alternatives in choice task *t* can be expressed as:

$$V_{ijt} = \mathbf{X}_{ijt}\mathbf{b}_i + p_{ijt}a_i + e_{ijt}\,, \qquad (1)$$

where $\mathbf{X}$ represents a vector of alternative-specific attributes, p is an additively separable cost, $\mathbf{b}$ and α are coefficients. Note that the coefficients are indexed by respondents – respondents' coefficients can differ and are assumed to follow *a priori* specified multivariate parametric distributions.

The stochastic component *e* allows for unobservable factors that affect individuals' choices. It has an unknown, possibly heteroskedastic variance $\left(\mathrm{var}\left(e_{ijt}\right) = s_i^2\right)$. The model is usually identified by normalizing this variance, making the error term $\varepsilon_{ijt} = e_{ijt}\cdot\sigma_i$, where $\sigma_i = \pi\Big/\left(\sqrt{6}s_i\right)$, identically and independently, extreme value type 1 distributed with a constant variance $\mathrm{var}\left(\varepsilon_{ijt}\right) = \pi^2/6$. This specification of the error term leads to convenient expression of choice probabilities – an individual will choose alternative *j* if $U_{ijt} > U_{ikt}$, for all $k \neq j$, and the probability that alternative *j* is chosen from a set of *J* alternatives becomes

$$P\left(j|J\right) = \frac{\exp\left(\boldsymbol{X}_{ijt}\left(\sigma_i\mathbf{b}_i\right) + p_{ijt}\left(\sigma_i a_i\right)\right)}{\sum_{k=1}^{J}\exp\left(\boldsymbol{X}_{ikt}\left(\sigma_i\mathbf{b}_i\right) + p_{ikt}\left(\sigma_i a_i\right)\right)}\,. \qquad (2)$$

Note that in the above specification, as a result of normalization the preference parameters became $\sigma_i\mathbf{b}_i$ and $\sigma_i a_i$. Due to the ordinal nature of utility (the preference coefficients do not have direct interpretation anyway), this specification still represents the same preferences for individual i.

Given that it is usually of interest to estimate WTP for the non-monetary attributes $\mathbf{X}$, it is convenient to introduce a modification which is equivalent to using a money-metric utility function (estimating preference parameters in WTP space, Train and Weeks 2005):

$$U_{ijt} = \sigma_i a_i\left(\boldsymbol{X}_{ijt}\frac{\sigma_i\mathbf{b}_i}{\sigma_i a_i} + p_{ijt}\right) + \varepsilon_{ijt} = \sigma_i a_i\left(\boldsymbol{X}_{ijt}\boldsymbol{\beta}_i + p_{ijt}\right) + \varepsilon_{ijt}\,. \qquad (3)$$

In this specification, the estimates obtained by a researcher are (1) a product of the scale and marginal utility of income $\sigma_i a_i$ and (2) the scale-free coefficients $\boldsymbol{\beta}_i$ corresponding to each of the choice attributes $\mathbf{X}$, which can be readily interpreted as respondents' marginal WTP for them.[14]

There exists no closed form expression of (2) when the coefficients are assumed random variables following the specified probability distributions. Instead, it can be simulated by averaging over *D* draws from the assumed distributions (Revelt and Train 1998). As a result, the simulated log-likelihood function becomes:

$$\log L = \sum_{i=1}^{N} \log \frac{1}{D} \sum_{d=1}^{D} \prod_{t=1}^{T_i} \frac{\exp\left(\sigma_i a_i \left(\boldsymbol{X}_{ijt}\boldsymbol{\beta}_i + p_{ijt}\right)\right)}{\sum_{k=1}^{J} \exp\left(\sigma_i a_i \left(\boldsymbol{X}_{ikt}\boldsymbol{\beta}_i + p_{ikt}\right)\right)} \quad . \tag{4}$$

Maximizing the simulated log-likelihood function in (4) allows to derive coefficient estimates, while the inverse of the negative of the Hessian at convergence becomes the approximation of the asymptotic variance-covariance matrix, allowing for deriving the standard errors associated with model estimates.

This model is the econometric basis of majority of SP studies, as its flexibility allows for incorporating various behavioral and respondent-specific factors in consumers' preferences.

3.4. Econometric approaches: multiple discrete-continuous choice models

Since the seminal paper of discrete choice models have become a state of the practice tool for analyzing individuals' preferences. Nevertheless, many decision of consumers are not strictly discrete choice, but rather a combination of discrete and continuous decisions. For example, individuals choose the places they will use for recreation (a discrete decision) and the number of times they will visit each recreational site (a continuous decision). In the literature, the different approach to modelling choice data of such nature was proposed, called multiple discrete-continuous choice models. In what follows, we employ Bhat (2005) and Bhat (2008) formulation, as the standard multinomial logit model arises as a special case of this specification. We change the notation slightly, to keep it consistent with the specification of discrete choice model described in the previous sub-section.

In the discrete-continuous setting we assume that individual *i* chooses the bundle of *J* goods $\mathbf{T_i} = \left(T_{1i}, T_{2i}, \ldots, T_{Ji}\right)$, where $T_{ji} \geq 0$ denotes the number of units of good *j* that individual consumes. Specifically, if $T_{ji} = 0$, then good *j* is not consumed. The utility individual derives from consuming this bundle is assumed to be

$$U\left(\mathbf{T_i}\right) = \sum_{j=1}^{J} \frac{\gamma_j}{\alpha_j} \exp\left(\mathbf{X_{ij}}\boldsymbol{\beta} + \varepsilon_{ij}\right) \left[\left(\frac{T_{ji}}{\gamma_j} + 1\right)^{\alpha_j} - 1\right], \tag{5}$$

where γ_j, α_j and $\boldsymbol{\beta}$ are parameters to be estimated, whereas ε_{ij} is a stochastic term, analogously as in the discrete choice setting. $\mathbf{X_{ij}}$ is a vector of explanatory variables, which may include attributes of different alternatives as well as socio-demographic characteristics. Note that in order to stay consistent with Bhat (2008) we assume that there is only one choice situation per individual and there is no preference heterogeneity. Nevertheless, these extensions can be easily incorporated into that framework, analogously as in the regular discrete choice models. The additive functional form of (5) implies that *J* goods considered are substitutes, but the more general utility specification can be employed, as in Vasquez Lavin and Hanemann (2008).

[14] Recall that calculating marginal WTP (implicit price) of an alternative requires calculating the marginal rate of substitution of the commodity for a monetary attribute. In the case utility function coefficients are known this becomes their ratio: b_i / a_i. If the model follows the specification described in (3), the ratio is: $\frac{\left(\sigma_i a_i\right) \cdot \beta_i}{\sigma_i a_i} = \beta_i$ and hence the coefficients associated with non-monetary attributes can be directly interpreted as marginal WTP (in the unit in which *p* is specified).

It is assumed that individual *i*, maximizes (5) with respect to $\sum p_{ji}T_{ji} = E_i$, where by E_i we denote total expenditures (for example, income) and p_{ji} is a price of good *j*. The model can be further extended to incorporate multiple constraints, as in Castro et al. (2012). If we denote by $e_{ji} = p_{ji}T_{ji}$ the expenditures on good *j*, and we denote by e_{ji}^* the optimal expenditures, we can then write down the probability of choosing the given bundle as

$$P\left(e_{1i}^*, e_{2i}^*, \ldots, e_{M_i i}^*, 0, \ldots, 0\right) = \frac{1}{\sigma^{M-1}}\left[\prod_{k=1}^{M_i} c_{ki}\right]\left[\sum_{k=1}^{M_i}\frac{1}{c_{ki}}\right]\left[\frac{\prod_{k=1}^{M_i}\exp\left(V_{ki}/\sigma\right)}{\left(\sum_{k=1}^{M_i}\exp\left(V_{ki}/\sigma\right)\right)^{M}}\right]\left(M_i - 1\right)!, \qquad (6)$$

where it is assumed that individual *i* consumes M_i first goods[15], $c_{ki} = \frac{1-\alpha_k}{e_{ki}^* + \gamma_k p_{ki}}$ and $V_{ki} = \mathbf{X_{ik}}\boldsymbol{\beta} + \left(\alpha_k - 1\right)\log\left(\frac{e_{ki}^*}{\gamma_k p_{ki}} + 1\right) - \log\left(p_{ki}\right)$. The probability in (6) is calculated assuming that error terms, ε_{ji}, have independent extreme value distributions with the scale of σ, which, contrary to the regular discrete choice models, can be estimated as long as there is variation in prices between alternatives.

This model is a substantial extension with respect to the simple MXL, as it allows to incorporate joint discrete-continuous choice processes, in which consumer chooses not only which goods to consume (possibly more than one), but also the amount of each of them (quantity). As a result, it may prove more appropriate in some choice settings, and hence avoid biases resulting from using inadequate model formulation.

4. Significance of the project

4.1. Significance and impact of the project

The debate over the validity of SP methods has once been very lively, particularly when fueled by policy-relevant controversies, such as the estimation of the environmental losses due to the Exxon Valdez oil spill in 1989 (Hausman 1993) and led to preparing a set of guidelines, which provided a framework for the estimation of valid SP responses (Arrow et al. 1993). The more recent **Deepwater Horizon oil spill** in 2010 renewed the interest in studying the validity of SP methods (Carson 2012b, Hausman 2012, Kling et al. 2012, Journal od Economic Perspectives), with high-profile publications such as Bishop et al. (2017, Science) being counter-attacked by industry-sponsored collections of problems associated with SP methods (McFadden and Train 2017). Overall, the **policy-interest again reignites the interest in studying the validity of SP methods**, and offers hope for a new, possibly updated set of guidelines which would allow SP-based welfare change estimates to be taken at face-value (c.f., Johnston et al. 2017).

Simply put, our research project intends to engage in the research stream regarding the **validity of SP methods** with respect to their **resilience to various behavioral anomalies**. Overall, this project fits perfectly into the current debate regarding the validity of SP methods and offers a chance to provide very relevant and timely contributions that **may even be of interest for the absolute top scientific journals**.[16]

Preference modeling underlies modern microeconomics, and it is difficult to overemphasize its importance. It is hard to overestimate the central role of preference modeling, and the SP methods in particular, in modern

[15] This assumption is not restricting, and it is used only to simplify the notation. We can simply assume that different individuals choose from the same number of goods, but they are indexed differently for each individual.

[16] Note the journals in which the articles referenced in the previews paragraph were published, such as *Science* or *Journal of Economic Perspectives*, and *Journal of Public Economics*. The researchers engaged in this research in the past include Nobel Prize laureates: Kenneth Arrow, Robert Solow or Daniel McFadden.

welfare economics and policy. These methods have been implemented in a vast number of papers in many fields of applied microeconomics, including marketing, transport, health and environmental economics (e.g., Ben-Akiva and Lerman 1985, Anderson et al. 1992, Haab and McConnell 2003, Hensher et al. 2005b, Louviere et al. 2006, Train 2009).

The practical significance of the project is very straightforward. By forming research-based conclusions and guidelines for the design of future SP studies we can lead to **improvements in methodology** that may have **substantial implications for policy analysis**. Our findings and recommendations regarding the construction of SP surveys and data analysis will have an effect on the quality of the so widely used SP-based decisions taken by the agencies, making the new policies better matched to the needs of the society.

Pragmatically, the project will allow us to make **several highly-significant contributions to the literature**. The principal investigator of the project has considerable experience (and a few dozen publications) with theoretical, methodological and applied aspects of modeling consumers' preferences. We are convinced, that the ideas described in this project proposal fill the gap in the existing knowledge and give a great possibility to contribute to the state-of-the-art literature – the implementation of the project will lead to **at least 6 highly-visible and widely-cited scientific publications** in the best journals in the field, such as *Journal of Environmental Economics and Management*, *Journal of Choice Modelling*, *Journal of Consumer Research*, *Environmental and Resource Economics*, *Ecological Economics*, *Land Economics*, *American Journal of Agricultural Economics*, *Health Economics*, *Transportation Research* or *Marketing Science*. In addition, we intend to disseminate our results in other ways, such as presentations at the most prestigious international conferences concerned with methodological aspects of choice modeling (e.g., *World Congress of Environmental and Resource Economics*, *International Choice Modeling Conference*, *Transportation Research Board Annual Meeting*, *World Congress of the International Health Economics Association*). Finally, we believe that making the software codes developed within the project available to other researchers and practitioners will substantially increase the scientific impact of our project and the visibility of our results, leading to a wider dissemination of the results (and more citations of our publications).

4.2. Pioneering nature of the project

The novel research program undertaken in this project will answer important methodological questions related to the behavioural anomalies observed in SP studies. Existing research in this area is inconclusive, as existing studies typically fail to control for the possible reasons for the observed anomalies attributing them instead to various behavioral phenomena. By engaging in the research outlined in this project, we will be able to fill this gap. We will address the research questions in a comprehensive way, while controlling for the other possible factors, which could influence the results. The project has a potential to be **ground-breaking** (for social science research standards), in the sense that it can **change the common perception** and, gradually, **the state-of-practice**. By systematically investigating the problems with the current methods and techniques, and providing solutions, we have a chance to make SP methods better, and hence improve validity and robustness of SP-based preference modelling and valuation.

4.3. Feasibility of research objectives

The research plan proposed in this project addresses an important and highly demanded aspect of consumers' preference elicitation. We have **identified the research gaps** in the state-of-the-art literature, related to the interpretation of the observed anomalies in SP studies, and proposed a **research program** that aims to investigate and propose solutions to these problems, thus making SP methods more **reliable** and the resulting welfare estimates **robust**. So far, state-of-the-art academic literature does not provide such research, although **there is a growing interest in such results**, particularly because of the potential importance of such findings for applied studies, and hence its policy relevance.

The principal investigator of this project has **substantial experience** in applying and developing the methodology of modeling consumers' preferences, as well as **publishing** in (nearly exclusively) top international journals, and in international cooperation, and **management of research projects**, some of which very large. Our team has the necessary knowledge and skills to warrant a successful implementation of the research plan proposed in this project. As a team, we believe we have a very good overview of the state-of-the-art literature and research trends in this field and the topic we identified and propose to research in this project fills the gap in the existing knowledge and will provide a novel contribution to the state-of-the-art literature.

5. Rationale for the establishment of a new research team

I have nearly completed my previous project (Sonata) that proved to be very successful in terms of its results and impact (nearly 10 scientific articles published in top journals in the field) and allowed for attracting very successful young researchers to this research area and to academia generally. These young researchers now celebrate successes of their own, successfully publishing, winning awards, scholarships and funding for their own research ideas. They still closely collaborate with me, each other, and other scientists in the field (including internationally).

I would like to use the opportunity this new project (Sonata Bis) provides to create a new, extended team of young researchers interested in the field of microeconomics and microeconometric analysis, valuation and modeling of consumers' preferences and behavioral economics. Such a team, backed up with the necessary infrastructure provided by the University of Warsaw, could successfully engage in solving state-of-the-art methodological problems and conduct high quality empirical analyses and in the future. As a result, we could become one of the leading centers devoted to the analysis of consumer choice in the world. I believe we are getting there already, and this project reflects what I think the best next steps for our future research and development are.

The project will provide four 30-month scholarships for PhD students (two scholarships for the first half of the project, and two scholarships for the second half of the project). It is expected, that after finishing the 30-month scholarship period each PhD student will stay in this or similar area of research and pursue their own high-quality research ideas, while being funded from scholarships for the best PhD students by the department, own research projects dedicated for PhD students (such as Preludium) or get involved as contributors in other researchers' projects. The project is planned for 5 years (2 x 2.5 years), as the research program is relatively work-intensive and combined of many interconnected components that need to be investigated jointly.

The estimated cost of co-investigators' salaries (scholarships) allow for attracting and involving the very best candidates, and making sure they do not need to participate in other (possibly less ambitious) applied projects to provide for themselves, while being involved in this research project. There are many exceptionally talented students of economics (including at the Department of Economics of the University of Warsaw), but most of them choose non-academia careers, particularly because of the insecurity associated with the funding of PhD studies. By offering scholarships I hope to change that and attract the smartest and the most hard-working students to PhD studies, to academia, and to my research group in particular.

The implementation of the project and the research programme, apart from leading to acquiring knowledge, will result in creating an integrated team, which will constitute a strong research centre dedicated to this important and dynamically developing area of economics. This team will be able to boldly take on new research problems, while cooperating with similar centres abroad, for example through international research projects and works on expert opinions. In addition, we intend to ensure the visibility of our research centre in Poland and abroad, where public administration may be interested in our research and expertise on modelling consumer preferences and valuation of nonmarket goods.

6. References

.

Ahn, J., G. Jeong, and Y. Kim. 2008. A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. Energy Economics **30**:2091-2104.

Ahtiainen, H., J. Artell, M. Czajkowski, B. Hasler, L. Hasselström, A. Huhtala, J. Meyerhoff, J. C. R. Smart, T. Söderqvist, M. H. Alemu, D. Angeli, K. Dahlbo, V. Fleming-Lehtinen, K. Hyytiäinen, A. Karlõševa, Y. Khaleeva, M. Maar, L. Martinsen, T. Nõmmann, K. Pakalniete, I. Oskolokaite, and D. Semeniene. 2014. Benefits of meeting nutrient reduction targets for the Baltic Sea – a contingent valuation study in the nine coastal states. Journal of Environmental Economics and Policy **3**:1-28.

Anderson, S. P., A. d. d. Palma, and J.-F. Thisse. 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

Aravena, C., P. Martinsson, and R. Scarpa. 2014. Does Money Talk? The Effect of a Monetary Attribute on the Marginal Values in a Choice Experiment. Energy Economics **44**:483-491.

Ariely, D., G. Loewenstein, and D. Prelec. 2003. "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences. Quarterly Journal of Economics **118**:73-105.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. Federal Register **58**:4601-4614.

Atkinson, G., B. Groom, N. Hanley, and S. Mourato. 2018. Environmental Valuation and Benefit-Cost Analysis in U.K. Policy. Journal of Benefit-Cost Analysis **9**:97-119.

Azevedo, C. D., J. A. Herriges, and C. L. Kling. 2003. Combining Revealed and Stated Preferences: Consistency Tests and Their Interpretations. American Journal of Agricultural Economics **85**:525-537.

Banzhaf, H. S., and V. K. Smith. 2007. Meta-analysis in model implementation: choice sets and the valuation of air quality improvements. **22**:1013-1031.

Bartczak, A., S. Chilton, M. Czajkowski, and J. Meyerhoff. 2017. Gain and loss of money in a choice experiment. The impact of financial loss aversion and risk preferences on willingness to pay to avoid renewable energy externalities. Energy Economics **65**:326–334.

Bateman, I. J., I. H. Langford, and J. Rasbash. 2001. Willingness-To-Pay Question Format Effects in Contingent Valuation Studies.*in* I. J. Bateman and K. G. Willis, editors. Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and developing Countries. Oxford University Press, Oxford, UK.

Ben-Akiva, M., and S. R. Lerman. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Bhat, C. R. 2005. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. Transportation Research Part B: Methodological **39**:679-707.

Bhat, C. R. 2008. The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. Transportation Research Part B: Methodological **42**:274-303.

Bhat, C. R., and S. Castelar. 2002. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. Transportation Research Part B: Methodological **36**:593-616.

Bhat, C. R., and S. Sen. 2006. Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. Transportation Research Part B: Methodological **40**:35-53.

Bishop, R. C., K. J. Boyle, R. T. Carson, D. Chapman, W. M. Hanemann, B. Kanninen, R. J. Kopp, J. A. Krosnick, J. List, N. Meade, R. Paterson, S. Presser, V. K. Smith, R. Tourangeau, M. Welsh, J. M. Wooldridge, M. DeBell, C. Donovan, M. Konopka, and N. Scherer. 2017. Putting a value on injuries to natural assets: The BP oil spill. Science **356**:253-254.

Blamey, R. K., J. W. Bennett, and M. D. Morrison. 1999. Yea-Saying in Contingent Valuation Surveys. Land Economics **75**:126-141.

Börger, T. 2016. Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment. Environmental and Resource Economics **65**:389-413.

Boyce, R. R., T. C. Brown, G. H. McClelland, G. L. Peterson, and W. D. Schulze. 1992. An Experimental Examination of Intrinsic Values as a Source for the WTA-WTP disparity. The American Economic Review **82**.

Boyle, K. J., H. F. MacDonald, H.-t. Cheng, and D. W. McCollum. 1998. Bid design and yea saying in single-bounded, dichotomous-choice questions. Land Economics **74**:49-64.

Brown, T. C., D. Kingsley, G. L. Peterson, N. E. Flores, A. Clarke, and A. Birjulin. 2008. Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement. Journal of Public Economics **92**:1595-1606.

Budziński, W. 2015. Znaczenie zmienności krańcowej użyteczności kosztu ponoszonego przez konsumenta dla wyceny dóbr nierynkowych. Ekonomia **43**:7-39.

Burrows, J., P. Dixon, and H. M. Chan. 2017. Response to Cost Prompts in Stated Preference Valuation of Environmental Goods. Pages 1-16 *in* D. McFadden and K. Train, editors. Contingent Valuation of Environmental Goods. A Comprehensive Critique. Edward Elgar Publishing, Northampton, MA.

Campbell, D., D. A. Hensher, and R. Scarpa. 2012. Cost thresholds, cut-offs and sensitivities in stated choice analysis: Identification and implications. Resource and Energy Economics **34**:396-411.

Caputo, V., E. J. Van Loo, R. Scarpa, J. R. M. Nayga, and V. W. 2018. Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments. Journal of Agricultural Economics **68**:35-57.

Carlsson, F., M. Kataria, and E. Lampi. 2016. Old and new aspects of respondent behavior in stated preference surveys. Paper presented at the Discrete Choice Modelling Workshop, Warsaw.

Carlsson, F., and P. Martinsson. 2008. How Much is Too Much? Environmental and Resource Economics **40**:165-176.

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

Carson, R. T. 2012a. Contingent Valuation: A Comprehensive Bibliography and History. Edward Elgar Publishing.

Carson, R. T. 2012b. Contingent Valuation: A Practical Alternative When Prices Aren't Available. Journal of Economic Perspectives **26**:27-42.

Carson, R. T., and M. Czajkowski. 2014. The Discrete Choice Experiment Approach to Environmental Contingent Valuation.*in* S. Hess and A. Daly, editors. Handbook of choice modelling. Edward Elgar, Northampton, MA.

Carson, R. T., and W. M. Hanemann. 2005. Contingent Valuation.*in* K. G. Mäler and J. R. Vincent, editors. Handbook of Environmental Economics. Elsevier, Amsterdam.

Castro, M., C. R. Bhat, R. M. Pendyala, and S. R. Jara-Díaz. 2012. Accommodating multiple constraints in the multiple discrete–continuous extreme value (MDCEV) choice model. Transportation Research Part B: Methodological **46**:729-743.

Champ, P. A., K. J. Boyle, and T. C. Brown. 2017. A Primer on Nonmarket Valuation. Springer, Amsterdam.

Cooper, J., and J. J. L. e. Loomis. 1992. Sensitivity of willingness-to-pay estimates to bid design in dichotomous choice contingent valuation models.211-224.

Cooper, J. C. 1993. Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys. Journal of Environmental Economics and Management **24**:25-40.

Coursey, D. L., J. L. Hovis, and W. D. Schulze. 1987. The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value. Quarterly Journal of Economics **102**:679-690.

Czajkowski, M., H. Ahtiainen, J. Artell, W. Budziński, B. Hasler, L. Hasselström, J. Meyerhoff, T. Nõmmann, D. Semeniene, and T. Söderqvist. 2015. Valuing the commons: An international study on the recreational benefits of the Baltic Sea. Journal of Environmental Management **156**:209-217.

Czajkowski, M., C. A. Vossler, W. Budziński, A. Wiśniewska, and E. Zawojska. 2017. Addressing empirical challenges related to the incentive compatibility of stated preference methods. Journal of Economic Behavior and Organization **142**:47-63.

Daly, A. 2010. Cost Damping in Travel Demand Models. RAND Corporation.
Day, B., and P. J.-L. Pinto. 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management **59**:271-285.
Figuerola, C. M. T., S. Colombo, and N. J. E. a. y. r. n. Hanley. 2014. Incorrectly accounting for preference heterogeneity in choice experiments: Implications for welfare measurement. **14**:97-121.
Fox, J., A. Daly, and B. Patruni. 2009. Improving the treatment of cost in large scale models.*in* European Transport Conference.
Freeman, A. M., J. A. Herriges, and C. L. Kling. 2014. The measurement of environmental and resource values: theory and methods. Routledge.
Glenk, K., J. Meyerhoff, F. Akaichi, and J. Martin-Ortega. forthcoming. Revisiting Cost Vector Effects in Discrete Choice Experiments. Resource and Energy Economics.
Griffiths, C., H. Klemick, M. Massey, C. Moore, S. Newbold, D. Simpson, P. Walsh, and W. Wheeler. 2012. U.S. Environmental Protection Agency Valuation of Surface Water Quality Improvements. Review of Environmental Economics and Policy **6**:130-146.
Gunn, H. 2001. Spatial and temporal transferability of relationships between travel demand, trip cost and travel time. Transportation Research Part E: Logistics and Transportation Review **37**:163-189.
Haab, T., and K. McConnell. 2003. The Econometrics Of Non-Market Valuation. Edward Elgar, Northampton, MA.
Hanley, N., W. Adamowicz, and R. E. Wright. 2005. Price vector effects in choice experiments: an empirical test. Resource and Energy Economics **27**:227-234.
Hanley, N., and M. Czajkowski. forthcoming. Stated Preference valuation methods: an evolving tool for understanding choices and informing policy. Review of Environmental Economics and Policy.
Hausman, J. A. 1993. Contingent Valuation: A Critical Assessment. North-Holland, Amsterdam.
Hausman, J. A. 2012. Contingent Valuation: From Dubious to Hopeless. Journal of Economic Perspectives **26**: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**: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., and M. J. M. L. Bradley. 1993. Using stated response choice data to enrich revealed preference discrete choice models. **4**:139-151.
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**:413-426.
Hensher, D. A., J. Rose, and W. H. Greene. 2005a. 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. 2005b. Applied Choice Analysis: A Primer. Cambridge University Press, Cambridge.
Hensher, D. A., J. M. Rose, and W. H. Greene. 2008. Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. Journal of Transport Geography **16**:126-133.
Hess, S., and N. Beharry-Borg. 2012. Accounting for Latent Attitudes in Willingness-to-Pay Studies: The Case of Coastal Water Quality Improvements in Tobago. Environmental and Resource Economics **52**:109-131.
Hess, S., and D. A. Hensher. 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B: Methodological **44**:781-790.
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 Behavior & Organization **96**:21-31.
Holmes, T. P., and R. A. Kramer. 2002. An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous-Choice Contingent Valuation. Journal of Environmental Economics and Management **29**:121-132.
Johnston, R. J., K. J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T. A. Cameron, W. M. Hanemann, N. Hanley, M. Ryan, R. Scarpa, R. Tourangeau, and C. A. Vossler. 2017. Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists **4**:319-405.
Jones, M. S., L. A. House, and Z. Gao. 2015. Respondent Screening and Revealed Preference Axioms: Testing Quarantining Methods for Enhanced Data Quality in Web Panel Surveys. Public Opinion Quarterly **79**:687-709.
Kling, C. L., D. J. Phaneuf, and J. Zhao. 2012. From Exxon to BP: Has Some Number Become Better than No Number? The Journal of Economic Perspectives **26**:3-26.
Knetsch, J. L., and J. A. Sinden. 1984. Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value. The Quarterly Journal of Economics:507-521.
Knittel, C. R., and K. Metaxoglou. 2014. Estimation of random-coefficient demand models: Two empiricists' perspective. Review of Economics and Statistics **96**:34-59.
Kragt, M. E. 2013. The Effects of Changing Cost Vectors on Choices and Scale Heterogeneity. Environmental and Resource Economics **54**:201-221.
Ladenburg, J., and S. B. Olsen. 2008. Gender-specific starting point bias in choice experiments: Evidence from an empirical study. Journal of Environmental Economics and Management **56**:275-285.
Lancaster, K. 1966. A New Approach to Consumer Theory. Journal of Political Economy **74**:132-157.
LaRiviere, J., M. Czajkowski, N. Hanley, M. Aanesen, J. Falk-Petersen, and D. Tinch. 2014. The Value of Familiarity: Effects of Experience, Knowledge and Signals on Willingness to Pay for a Public Good. University of Stirling Working paper.
Li, L., W. Adamowicz, and J. Swait. 2015. The effect of choice set misspecification on welfare measures in random utility models. Resource and Energy Economics **42**:71-92.
Louviere, J. J., D. A. Hensher, and J. D. Swait. 2006. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.
Luisetti, T., I. J. Bateman, and R. K. J. L. E. Turner. 2011. Testing the fundamental assumption of choice experiments: are values absolute or relative? **87**:284-296.
Mabit, S. L., J. Rich, P. Burge, and D. Potoglou. 2013. Valuation of travel time for international long-distance travel – results from the Fehmarn Belt stated choice experiment. Journal of Transport Geography **33**:153-161.

McFadden, D., and K. Train. 2017. Contingent Valuation of Environmental Goods. A Comprehensive Critique. Edward Elgar Publishing, Northampton, MA.
Menegaki, A. N., S. B. Olsen, and K. P. Tsagarakis. 2016. Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. Journal of Choice Modelling **18**:18-50.
Meyerhoff, J., and K. Glenk. 2015. Learning how to choose—effects of instructional choice sets in discrete choice experiments. Resource and Energy Economics **41**:122-142.
Mørkbak, M. R., T. Christensen, and D. Gyrd-Hansen. 2010. Choke Price Bias in Choice Experiments. Environmental and Resource Economics **45**:537-551.
Parsons, G., and K. Myers. 2017. Fat tails and truncated bids in contingent valuation: an application to an endangered shorebird species.*in* D. McFadden and K. Train, editors. Contingent valuation of environmental goods. A Comprehensive Critique. Edward Elgar, Northampton, USA.
Parsons, G. R. 2017. The Travel Cost Model. Pages 187-233 *in* P. A. Champ, K. J. Boyle, and T. C. Brown, editors. A Primer on Nonmarket Valuation. Springer Netherlands.
Revelt, D., and K. Train. 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Review of Economics and Statistics **80**:647-657.
Ryan, M., and S. Wordsworth. 2000. Sensitivity of Willingness to Pay Estimates to the Level of Attributes in Discrete Choice Experiments. **47**:504-524.
Sandorf, E. D., D. Campbell, and N. Hanley. forthcoming. Disentangling the influence of knowledge on attribute non-attendance. Journal of Choice Modelling.
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**:165-180.
Stathopoulos, A., and S. Hess. 2012. Revisiting reference point formation, gains–losses asymmetry and non-linear sensitivities with an emphasis on attribute specific treatment. Transportation Research Part A: Policy and Practice **46**:1673-1689.
Su, L., B. D. Adam, J. L. Lusk, F. J. J. o. a. Arthur, and r. economics. 2017. Anchoring, Information, and Fragility of Choice Experiments: An Application to Consumer Willingness to Pay for Rice with Improved Storage Management. Journal of Agricultural and Resource Economics **42**:255.
Sugden, R., J. Zheng, and D. J. Zizzo. 2013. Not all anchors are created equal. Journal of Economic Psychology **39**:21-31.
Svenningsen, L. S., and J. B. Jacobsen. 2018. Testing the effect of changes in elicitation format, payment vehicle and bid range on the hypothetical bias for moral goods. Journal of Choice Modelling.
Swait, J. 2001. A non-compensatory choice model incorporating attribute cutoffs. Transportation Research Part B: Methodological **35**:903-928.
Swait, J., and W. Adamowicz. 2001. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. The Journal of Consumer Research **28**:135-148.
Taylor, L. O. 2017. Hedonics. Pages 235-292 *in* P. A. Champ, K. J. Boyle, and T. C. Brown, editors. A Primer on Nonmarket Valuation. Springer, Amsterdam.
Termansen, M., C. J. McClean, and F. S. Jensen. 2013. Modelling and mapping spatial heterogeneity in forest recreation services. Ecological Economics **92**:48-57.
Thiene, M., J. Swait, and R. Scarpa. 2017. Choice set formation for outdoor destinations: The role of motivations and preference discrimination in site selection for the management of public expenditures on protected areas. Journal of Environmental Economics and Management **81**:152-173.
Train, K. E. 2009. Discrete Choice Methods with Simulation. 2 edition. Cambridge University Press, New York.
Train, K. E., and M. Weeks. 2005. Discrete choice models in preference space and williingness-to-pay space.*in* R. Scarpa and A. Alberini, editors. Applications of simulations methods in environmental and resource economics. Springer, Dordrecht.
US Environmental Protection Agency. 1994. President Clinton's clean water initiative: analysis of benefits and costs. EPA 800-R-94-002 (NTIS Document No. PB94-154101). Office of Water, United States Environmental Protection Agency, Washington DC.
Vasquez Lavin, F., and W. M. Hanemann. 2008. Functional forms in discrete/continuous choice models with general corner solution.
Veronesi, M., A. Alberini, and J. Cooper. 2011. Implications of Bid Design and Willingness-To-Pay Distribution for Starting Point Bias in Double-Bounded Dichotomous Choice Contingent Valuation Surveys. Environmental and Resource Economics **49**:199-215.
von Haefen, R. H. 2008. Latent Consideration Sets and Continuous Demand Systems. Environmental and Resource Economics **41**:363-379.
von Haefen, R. H., and D. J. Phaneuf. 2003. Estimating preferences for outdoor recreation:: a comparison of continuous and count data demand system frameworks. Journal of Environmental Economics and Management **45**:612-630.
von Haefen, R. H., D. J. Phaneuf, and G. R. Parsons. 2004. Estimation and Welfare Analysis With Large Demand Systems. Journal of Business and Economic Statistics **22**:194-205.
Vossler, C. A., M. Doyon, and D. Rondeau. 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. American Economic Journal: Microeconomics **4**:145-171.
Vossler, C. A., and J. S. Holladay. 2018. Alternative value elicitation formats in contingent valuation: Mechanism design and convergent validity. Journal of Public Economics **165**:133-145.
Whitehead, J. C., S. K. Pattanayak, G. L. Van Houtven, and B. R. Gelso. 2008. Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science. **22**:872-908.
Zawojska, E., and M. Czajkowski. 2017. Re-examining empirical evidence on stated preferences: Importance of incentive compatibility. Journal of Environmental Economics and Policy **6**:374-403.