How Sensitive are Environmental Values to Payment Card Design in Contingent Valuation?

Magnus Aagaard Skeie (PhD candidate[a]), Henrik Lindhjem (Partner[b]), Ståle Navrud (Professor[c]), Tobias Otterbring (Professor[a, d])

Corresponding author: Magnus Aagaard Skeie.

[a]University of Agder, School of Business and Law, Universitetsveien 19, NO-4630, Kristiansand S, Norway. Email: magnus.ivar.aagaard.skeie@uia.no

[b]Menon Economics, Sørkedalsveien 10B, NO-0369, Oslo, Norway. Email: henrik.lindhjem@menon.no

[c]Norwegian University of Life Sciences, School of Economics and Business, Elizabeth Stephansens vei 15, NO-1433, Ås, Norway. Email: stale.navrud@nmbu.no

[d] School of Electrical Engineering and Computer Science, Division of Media Technology and Interaction Design, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden. Email: tobias.otterbring@uia.no

Abstract

Contingent valuation studies of people's willingness-to-pay for ecosystem services are frequently used to inform the social benefit-cost analysis of environmental protection measures. While contingent valuation is generally accepted in this context, response anomalies exist. Drawing on the anchoring-and-adjustment heuristic in psychology and the context effects literature, we examine the impact of subtle design variation on people's willingness-to-pay. In a split-sample national survey of willingness-to-pay to prevent coastal environmental damages from oil spills, different payment card elicitation formats significantly

doi:10.3368/le.101.3.061924-0053R

affect mean willingness-to-pay. This underscores the importance of a research agenda on the effects of subtle design variation on environmental value estimates.

1. Introduction

The valuation of changes in ecosystem services is often based on stated preference studies that seek to quantify individuals' preferences in monetary terms (Jensen, Johnston, and Olsen 2019; Grant and Langpap 2020; Luther, Swinton, and Deynze 2022). For major industrial disasters, such as the Exxon Valdez oil spill in 1989 or the BP Deepwater Horizon oil spill in 2010, ecosystem service damages manifest as a combination of interdependent attributes. In such cases, the contingent valuation method has been used to estimate the welfare loss (Carson et al. 2003; Bishop et al. 2017).

Since the Exxon Valdez oil spill litigation (Carson et al. 1992), studies have extensively debated the validity of contingent valuation. Although the usefulness and validity of contingent valuation are now largely established (Kling, Phaneuf, and Zhao 2012), behavioral response anomalies prevail (Dietz, Stern, and Dan 2009; Lusk and Norwood 2009; Alevy, List, and Adamowicz 2011; Börger et al. 2021; Mahieu et al. 2024; Zawojska, Welling, and Sagebiel 2024). These anomalies are theoretically significant as they may indicate previously overlooked aspects of preference elicitation. Additionally, from a practical perspective, even minor variations in stated preferences caused by the contingent valuation survey design can result in significant cumulative effects on welfare estimates when aggregated across a large, affected population (Johnston et al. 2017). Therefore, an improved understanding of the influence of the contingent valuation survey design on ecosystem service valuation is of great practical importance for both government bodies and resource managers.

The influence of the payment elicitation format on stated preferences, so-called elicitation effects, is a key topic in debates over the validity of the contingent valuation method. The two main categories of explanations for elicitation effects include differences in perceived economic incentives across formats and behavioral reasons (Vossler and Zawojska 2020). The former necessitates optimizing the survey during the design phase to ensure economically motivated responses (Carson, Flores, and Meade 2001; Carson and Groves 2007). The latter may indicate a fundamental context dependence of preferences (Ariely, Loewenstein, and Prelec 2003; Zhang and Adamowicz 2011). Vossler and Zawojska (2020) failed to reject convergent validity in the willingness-to-pay (WTP) across several popular elicitation formats when controlling for economic incentives, suggesting that behavioral factors might not primarily drive elicitation effects across formats. Nonetheless, substantial evidence indicates that behavioral factors do produce elicitation effects within specific formats upon bid design variation. Notable examples include starting point effects in dichotomous choice formats (e.g., Holmes and Kramer 1995; Carson, Flores, and Meade 2001) and range effects in the payment card format (e.g., Covey, Loomes, and Bateman 2007; Soeteman, van Exel, and Bobinac 2017). If behavioral factors do play a role in bid design in a given elicitation format, even when incentive compatibility is established (as for the single-bounded dichotomous choice format; Johnston et al., 2017), it warrants attention in the study design phase.

In this study, we focus on the payment card elicitation format. It is recognized as a simple, cost-effective, and statistically efficient way to examine the sensitivity of preferences across design variations (Schläpfer and Getzner 2020), and is popular in contingent valuation research owing to its practical advantages (Voltaire et al. 2019). Moreover, this format was designed to mitigate the starting-point bias by presenting a set of bids rather than single bids on which respondents tend to anchor their WTP responses (Mitchell and Carson 1981), making it an interesting study object for bid design variation.

Our study is centered on the following research questions: Do contingent valuation respondents anchor the stated WTP on a subset of initial bids in the payment card elicitation format? If so, is the anchoring effect on WTP both statistically and economically significant? The first question pertains to the notion of preference construction in environmental valuation (Gregory, Lichtenstein, and Slovic 1993). This view holds that as people are not used to evaluating environmental goods in monetary terms, preferences are constructed in the elicitation process, aided by available contextual cues and information. Anchoring on a subset of initial payment card bids as a contextual cue, coupled with insufficient adjustment (Tversky and Kahneman 1974; Epley and Gilovich 2006), may yield systematic differences in the mean WTP for different bid designs. The second question pertains to whether any observed effects are not only statistically significant but also "large enough to matter for policy or science" (McCloskey and Ziliak 1996, p. 104), denoting economic or practical significance.

To address the research questions, we conducted a split-sample experiment within a large-scale contingent valuation survey focused on safety measures to prevent coastal oil spills in Norway. The respondents were randomly assigned to one of three payment card conditions, each differing in bid design, particularly in the first part of the bids. This design enabled causal inference of the anchoring effects of the bid design on the stated WTP. We use interval regression to test for such effects while controlling for other relevant covariates. Further, we assess the robustness of the results by excluding the fastest, and perhaps most inattentive, survey respondents (i.e., speeders; Sandorf, Aanesen, and Navrud 2016) and by re-estimating the interval regression model with Tobit regression. To evaluate the economic significance of our results, we demonstrate the resulting differences in aggregate welfare estimates used in social benefit-cost analysis at a national government agency.

Our findings indicate that the payment card bid design significantly influences the stated preferences. Higher initial bids significantly increase the predicted mean WTP from the interval regression model, at most by 43%. Further, kernel density plots show clear shifts in estimated WTP distributions across the experimental payment card conditions. Methodologically, our results suggest that payment card bids have a certain contextual influence that should be accommodated in contingent valuation study design to mitigate the influence of practitioners' own design choices on environmental value estimates. Practically, our research highlights how the accumulation of individual-level effects on stated preferences, as evident in our data, can impact the aggregate welfare estimates used in social benefit-cost analysis.

We contribute to research on elicitation effects in two focal ways. Our primary contribution is methodological, by testing for anchoring effects on stated preferences in an elicitation format specifically designed to mitigate such effects (Mitchell and Carson 1981). While numerous studies include payment cards in elicitation format comparisons (e.g., Ryan, Scott, and Donaldson 2004; Blaine et al. 2005; Chestnut, Rowe, and Breffle 2012), few studies have systematically investigated the effect of bid design variation on stated preferences within the payment card format alone (Rowe, Schulze, and Breffle 1996; Smith 2006; Soeteman, van Exel, and Bobinac 2017), which is a methodological issue facing all practitioners using this elicitation format. Our secondary contribution is practical, as our study is distinctly policy-oriented. By conducting a split-sample experiment within a contingent valuation survey designed to inform social benefit-cost analysis at a national government agency, we demonstrate the economic significance of an individual-level anchoring effect on stated preferences in a policy context. Additionally, we offer specific advice to practitioners on how to accommodate the contextual influence of the payment elicitation format on stated preferences within contingent valuation study designs to minimize the impact of the researcher's bid design choices on value estimates.

2. Conceptual Framework and Hypothesis Development

Context effects in stated preference literature

Many studies have reported on context effects in stated preference research. Faccioli and Glenk (2022) recently proposed a classification in which context effects are the broadest category of studies, subsuming information effects and, in turn, framing effects. They define context effects as "a set of factors of the decision environment that have the potential to shift the choice outcome by altering the process by which the decision is made" (p. 315). Examples of studies of context effects on stated preferences in the broadest sense include the effects of different preference elicitation formats (Vossler and Holladay 2018; Vossler and Zawojska 2020), different presentation orders of mutually exclusive goods (; Powe and Bateman 2003; Clark and Friesen 2008), different survey modes (e.g., Lindhjem and Navrud 2011), and different degrees of complexity in discrete choice experiments (Meyerhoff, Oehlmann, and Weller 2015; Akinc, Street, and Vandebroek 2024). This study focuses on context effects in the broadest sense.

The effects of the preference elicitation format on stated preferences are commonly denoted as elicitation effects in studies of context effects (Vossler and Zawojska 2020; Faccioli and Glenk 2022). Many studies have tested convergent validity in WTP across elicitation formats. Holmes and Kramer (1995) compared the WTP from the dichotomous choice and payment card formats in a contingent valuation study of forest ecosystem protection and reject convergent validity. This result has been corroborated by later studies (Ready, Navrud, and Dubourg 2001; Ryan, Scott, and Donaldson 2004; Champ and Bishop 2006). Further, Green et al. (1998) rejected convergent validity in WTP between the dichotomous choice and open-ended formats in the contexts of seabird protection and increased traffic safety. Welsh and Poe (1998) tested the convergent validity in their multiple-bounded discrete choice format with both the dichotomous choice, payment card, and open-ended formats. They found practically indistinguishable cumulative WTP distributions for the multiple-bounded, payment card, and open-ended formats depending on the coding of response certainty levels in the multiple-bounded format, and they rejected convergent validity in WTP by comparing either of these formats to the dichotomous choice format. Bateman et al. (1995) and Frew, Wolstenholme, and Whynes (2003) included the iterative bidding and the bidding game formats, respectively, in comparisons with several of the abovementioned formats. Both studies rejected convergent validity. However, Vossler and Zawojska (2020) recently found that when economic incentives are controlled for, by ensuring incentive compatibility in different elicitation formats, the expected elicitation effects commonly reported in the literature do not occur across four widely used formats.[1]

Beyond elicitation effects across different formats, context effects studies have also focused on the potential effects of bid design within formats. In this paper, the term bid design refers to the choice of a single amount (for the single-bounded dichotomous choice format) or a set of amounts (for other relevant formats) constituting the bid(s) presented to respondents. For the single-bounded dichotomous choice format, higher bids have been shown to yield higher WTP through “yea-saying” (Cooper and Loomis 1992; Holmes and Kramer 1995; Kanninen 1995; Boyle et al. 1998). The initial bid may also anchor subsequent responses in double- and multiple-bounded dichotomous choice formats (Herriges and Shogren 1996; Whitehead 2002; Veronesi, Alberini, and Cooper 2011). In multiple-bounded discrete choice formats (Welsh and Poe 1998), with polychotomous response options soliciting respondents’ voting certainty associated with each bid, presenting bids in descending, rather than ascending, order has been shown to increase WTP (Roach, Boyle, and Welsh 2002; Alberini, Boyle, and Welsh 2003;

Wang et al. 2013). Regarding bid design, the multiple-bounded discrete choice format most closely resembles the payment card format; however, these two formats differ in preference elicitation. Whereas the former solicits a series of responses, including voting certainty, the latter solicits a single response in the form of choosing a bid representing the maximum WTP (Welsh and Poe 1998). Therefore, elicitation effects may also differ.

Only a few studies have investigated how payment card bid design impacts WTP. One was in the environmental domain (Rowe, Schulze, and Breffle 1996), and two were in the health domain (Smith 2006; Soeteman, van Exel, and Bobinac 2017). First, Rowe, Schulze, and Breffle (1996) studied the effect of differing payment card ranges and center values on WTP for the cleanup of hazardous waste-sites in Colorado by using four payment card versions. They found no evidence of a range effect on WTP across versions 1–3, with endpoints of $1000, $5000, and $10,000, respectively. However, they later observed an effect with the fourth version, which cut the highest 20% of the observed WTP solicited by version 1-3 to set its endpoint. Here, the observed cumulative distribution of WTP is truncated from above, in turn yielding a lower mean WTP. The authors suggest that truncation of the WTP distribution in bid design, rather than differences in an arbitrary range exceeding the end of the WTP distribution, affect the elicited WTP.

Second, Smith (2006) studied the influence of the bid order (low-to-high, high-to-low, and random) on the stated WTP for health gains. He found that a high-to-low ordering of payment card bids yields a higher mean WTP than both low-to-high and random ordering. Third, Soeteman, van Exel, and Bobinac (2017) examined the impact of two payment card versions with varying starting points, midpoints, endpoints, and interval sizes on respondents' WTP for health gains. They found a higher mean WTP from respondents using the payment scale with the higher starting point, midpoint and endpoint. Nevertheless, the impacts of the payment card bid design on the stated WTP are not well understood.

Hypothesis development

The existence of elicitation effects may be indicative of some degree of preference construction (Payne, Bettman, and Johnson 1992; Slovic 1995; Lichtenstein and Slovic 2009) among contingent valuation respondents. The standard neoclassical economics view of preferences is based on what Fischhoff (1991) calls a "philosophy of articulated values," which can be placed on one end of a continuum of philosophies for preference construction. At this end, individuals have well-defined preferences for which mere consultation is required to assign value to a certain good (Kahneman et al. 1993). The opposite end of the continuum represents a "philosophy of basic values," which is the dominant view of value formation in psychology and decision making (Schwarz 1997). This position holds that individuals lack well-differentiated values for all but the most familiar of evaluation questions. Otherwise, individuals must derive values from their basic values through some inferential process (Fischhoff 1991). If preferences are constructed, at least to some degree, anchoring on available information cues when constructing the WTP response may be inevitable.

Anchoring, sometimes referred to as the anchoring-and-adjustment heuristic, originally refers to effects observed in studies of judgment and decision-making under uncertainty, where people use an initial value as an anchor (i.e., reference point) from which they adjust to arrive at a final answer (Tversky and Kahneman 1974). As the adjustment process is typically insufficient, different starting points tend to yield different mean estimates, biased toward the value of the anchor. This holds even with uninformative, arbitrary anchors. For instance, Tversky and Kahneman (1974) reported on a study where respondents were to estimate the percentage of African countries in the United Nations after receiving an anchor value determined by spinning a wheel of fortune from 0 to 100. The arbitrary anchors had a significant impact on the estimates, and it persisted through the introduction of payoffs for accuracy. In the context of preference elicitation, Ariely, Loewenstein, and Prelec (2003) demonstrated that the stated WTP may be anchored on arbitrary value ques for the initial elicitation (e.g., last two digits of respondents' social security number). At a more general level, anchoring has been shown to be substantial and robust across contexts and populations (Green et al. 1998).

Whether and how anchoring takes place in the payment card format remains unclear. Some researchers contend that the payment card is not prone to anchoring as there is no single initial bid appearing as an obvious anchor (Boyle 2003; Soeteman, van Exel, and Bobinac 2017), whereas others suggest possible anchoring on the whole payment card range (Dubourg, Jones-Lee, and Loomes 1997) or on the starting values (Chanel, Makhloufi, and Abu-Zaineh 2017). Beyond finding evidence of higher mean WTP with a high-to-low payment card bid sorting relative to low-to-high and random sorting alternatives, Smith (2006) found no evidence of differences in WTP between the latter two. This suggests that for the payment card format, the first, or first few, bids may serve as a more pronounced anchor than the endpoint (or range).

In this study, we use the anchoring-and-adjustment heuristic (Tversky and Kahneman 1974; Epley and Gilovich 2006) as a framework for understanding the results of our experimental variation in payment card bid design. Further, as no bid is presented in isolation, we posit that a set of bids from the first part of the payment card may collectively anchor the stated WTP. Therefore, stating the WTP using a payment card elicitation format with a set of higher (lower) bids in any given first part of the payment card should reasonably result in higher (lower) estimates of the mean WTP owing to anchoring and insufficient adjustment. Formally, we test the validity of the following main hypothesis in this paper:

> H1: Increasing the value of the bids on the first part of the payment card leads to an increase in the mean WTP.

We test our hypothesis in a split-sample experimental design. We randomly assign respondents to three versions of a payment card elicitation format in a contingent valuation survey eliciting preferences to avoid ecosystem service damage due to coastal oil spills from ship accidents in the Oslofjord region of Eastern Norway. Further details on our payment card format are provided later in Section 2, and the experimental design and contingent valuation survey are described in greater depth in Section 3.

Payment card operationalization

Conceptually, a conventional payment card (Mitchell and Carson 1981) is a set of bids that represent the potential costs of a proposed intervention. Respondents are asked to specify their maximum WTP from this set. Let the bids, arranged in ascending order, be denoted as c_1, c_2, …, c_N. If a respondent selects a bid $s = c_n$, it can be interpreted as a "yes" vote for the intervention at any cost within the range $[c_1, c_n]$ and a "no" vote for costs exceeding c_n (i.e., in the range $[c_{n+1}, c_N]$) (Vossler and Holladay 2018).

Each elicitation format has its advantages and disadvantages (see, e.g., Boyle 2003). For the payment card, Reaves, Kramer, and Holmes (1999) contend that it exhibits favorable characteristics in terms of survey response rates, item nonresponse, and protest bids. Additionally, the payment card format has consistently provided more conservative results in elicitation compared to other formats (e.g., Ryan, Scott, and Donaldson 2004; Champ and Bishop 2006). Nonetheless, the payment card format is prone to incentive incompatibility in preference elicitation (Vossler and Holladay 2018). First, presenting several bids may induce a sense of project cost uncertainty and lead respondents to believe they can influence the project cost by their own statement of WTP. Second, a clear implementation rule is generally lacking in terms of the link between stating one's WTP and its impact on the probability of project implementation. We discuss the incentive compatibility of our operationalization of the payment card format in Section 3.

The payment card format has been operationalized in several ways in the literature. Random card sorting was first introduced by Carthy et al. (1999) to address range bias. An interviewer shuffles a stack of cards, each with a single bid, and asks the respondent to sort each card into one of three piles: "definitely would pay," "unsure," and "definitely would not pay." Nonetheless, later studies showed that the random card sorting format suffers from range bias (Covey, Loomes, and Bateman 2007; Shackley and Dixon 2014). Another payment card variant to mitigate bias is the circular payment card (Chanel, Makhloufi, and Abu-Zaineh 2017; Champonnois, Chanel, and Makhloufi 2018), where the intervals implied by the conventional payment card bids are mapped onto a pie-chart. Visually, it eliminates the starting and end points, and respondents must spin the wheel to arrive at the interval containing their maximum WTP. Chanel, Makhloufi, and Abu-Zaineh (2017) found that the WTP was significantly higher with the circular compared to the conventional payment card, which they attribute to less anchoring on starting values. Other payment card variants incorporate respondent uncertainty even more explicitly than does random card sorting. The stochastic payment card (Wang 1997; Wang and Whittington 2005) asks respondents to indicate the percentage probability of being willing to pay each bid, within the bounds of the five-level ordinal uncertainty scale from the multiple-bounded discrete choice format (Welsh and Poe 1998). The polychotomous choice payment card (Ndambiri, Brouwer, and Mungatana 2016; Ndambiri, Mungatana, and Brouwer 2017) elicits uncertainty responses for each bid using only the abovementioned ordinal scale. However, none of the alternative variants of the conventional payment card seem to have been widely adopted in the contingent valuation literature.

In this study, we use a conventional payment card format as our point of departure. However, we have sought to ease the elicitation process in the online survey by putting the bids on a horizontal scale, where respondents indicate their WTP by placing the cursor (from its default position left of zero) on the bid that reflects their maximum WTP, as illustrated in Figure 1.[2] Our payment scale resembles the conventional payment card in all features but the visual appearance. This sort of payment scale has been applied in both health (Soeteman, van Exel, and Bobinac 2017) and environmental economics (Li et al. 2021). The far-right side includes an option to state a WTP "more than [the highest bid]" and a "don't know" option (see, e.g., Rowe, Schulze, and Breffle 1996). As online contingent valuation surveys are increasingly conducted on small-screen mobile devices[3], we believe it is of high value to subject more user-friendly payment card variants to tests of both preference elicitation itself and sensitivity in elicitation to bid design variation.

<<Fig 1 about here>>

3. Data and Methods

We use a three-way split sample design based on two simultaneous data collections to investigate anchoring effects in the payment scale. The first data collection pertained to a national contingent valuation survey to avoid ecosystem service damage due to oil spills in Norway. Preferences were elicited in five regions by using a payment scale as described in Section 2 to estimate associated welfare losses for use in social benefit-cost analysis. The second data collection pertained solely to an experimental study of payment card elicitation effects, and it elicited preferences from additional respondents drawn from the Eastern region of the national contingent valuation survey, with random assignment to two payment scales differing both from each other and that of the national survey. The surveys were identical across the two data collections up to the point of preference elicitation.

There were two differences in elicitation between the two survey versions, beyond the intended experimental variation. First, respondents in the experimental survey were allowed to state an open-ended WTP response after selecting a payment scale bid, but only within the interval implied by the selected bid and the next bid. This exploratory feature was included to obtain more accurate WTP data for the experimental study.[4] Second, while the national survey asked for household WTP, the experimental survey asked for individual WTP. While this difference is important for aggregate welfare estimates post elicitation, Lindhjem and Navrud (2009) found no significant difference in the elicitation itself between asking for individual and household WTP in split samples. Nonetheless, we discuss the implication of the difference in elicitation for our results in Section 5. Post elicitation, the national survey was more extensive than the experimental survey. Having access to both data sets enabled us to expand the experimental study with an additional payment scale condition.

The contingent valuation survey included scope programs and elicited WTP sequentially to avoid four levels of ecosystem service damage pertaining to four oil spills of increasing magnitude. Consequently, we elicit WTP in a 3 (payment scale: Control, Treatment 1, Treatment 2) between-subject × 4 (ecosystem service damage: small, medium, large, very large) within-subject design. To test for elicitation effects across payment scales, we use interval regression with a categorical variable for the payment scale condition and other covariates, estimated separately for each of the four ecosystem service damage levels. Subsequently, we re-estimate the models when speeders are excluded (Sandorf, Aanesen, and Navrud 2016) and with Tobit regression, respectively, as robustness checks. The following subsections provide more details on the study design.

Experimental conditions

To test our focal hypothesis, we randomly assigned the respondents to three payment scale conditions. The Control condition (depicted in panel a of Figure 1) displays progressively increasing bids, in line with established practices in the literature (Rowe, Schulze, and Breffle 1996; Champonnois, Chanel, and Makhloufi 2018), that encompass a range of plausible WTPs in the population. To uncover potential anchoring in the first part of the payment scale, we first amplify the differences in bid design in this part of the scale. Consequently, Treatment 1 (shown in panel b) features bids that increase linearly by NOK 100 until a midpoint, beyond which they increase by NOK 1,000. The difference in bids compared to the Control condition is ten-fold for the first (nonzero) bid and then four-, three-, and two-fold for the second, third and fourth bids, respectively. Treatment 2 (depicted in panel c) further increases bids in the first part of the scale by escalating bids by NOK 200 up to the midpoint. The increase in bids on the first half of the scale is only two-fold compared to that in Treatment 1; however, it effectively doubles the amplification of the abovementioned differences from the Control to the Treatment 1 condition.

The final bid of the payment scales slightly varies between the Control and Treatment conditions. Specifically, the Control condition ends at NOK 12,000 and the Treatment conditions end at NOK 10,000. This discrepancy results from trade-offs made during the design phase of the Treatment conditions to isolate potential anchoring effects in the first part of the scale Consequently, the number of bids and the effective midpoint and endpoints of the Treatment conditions were evaluated. Importantly, each payment scale also had to make stand-alone intuitive sense to the assigned respondents in terms of bid evolution. As this study focuses on potential anchoring in the first part of the payment scale, we wanted to amplify the differences across conditions in this part. The abovementioned trade-offs resulted in linearly increasing bids (Mitchell and Carson 1981; Champonnois, Chanel, and Makhloufi 2018) in the first half of the Treatment conditions. Potential limitations owing to the difference in endpoints between the Control and Treatment conditions are addressed in Section 5.

The contingent valuation question was identical for all four valuation scenarios, beyond the abovementioned difference in WTP and the insertion of the relevant damage level qualifier: "Please indicate in the figure [i.e., the payment scale] below the greatest amount, if any, that [you/your household] [are/is] definitely willing to pay as an earmarked one-time tax for new measures to prevent [small/medium/large/very large] environmental damage in the Oslofjord area." More details on the contingent valuation survey are provided later in Section 3.

Respondents

As we assumed small elicitation effects by conventional standards, we sought to obtain a sufficiently large sample size to be able to detect such effects with high statistical power. Accordingly, our goal was to have a minimum of 650 valid observations of WTP for each of the three conditions. This would yield 95% statistical power to detect mean differences between conditions corresponding to an effect size of $d = 0.20$ or $r = 0.10$, assuming the conventional alpha level of $\alpha = 0.05$ (Faul et al. 2007). Factoring in attrition due to outliers, inconsistent response patterns, and protest responses, we sought to recruit a minimum of 1000 respondents per condition. This target sample size was met in our original sample ($N = 3707$ and cell sizes ranging from $n = 1177$ to $n = 1272$). The sample is representative of the affected population; further details are provided later in Section 3.

Data cleaning and empirical strategy

The dataset comprises 1,272, 1,258, and 1,177 respondents in the Control, Treatment 1, and Treatment 2 conditions, respectively. Across conditions, 291–327 respondents were excluded during the pre-analysis based on commonly applied exclusion criteria in contingent valuation. These criteria include WTP outliers (WTP > 5% of income[5]), zero-responses with protest motives (e.g., Boyle 2003), and inconsistent responses (decreasing WTP with increasing environmental improvement; Veisten et al. 2004). After exclusions, the resulting subsamples contain 965, 931, and 886 respondents in the Control, Treatment 1, and Treatment 2 conditions, respectively. Detailed counts regarding the exclusions are provided in Table 1. In terms of balance, a 3 (experimental condition: Control, Treatment 1, Treatment 2) × 2 (excluded: yes, no) chi-square analysis indicates that attrition due to exclusion is not systematically different across conditions, $\chi^2(2, N = 3707) = 1.21, p = 0.55$. Further analyses showed that the conditions are balanced on central socioeconomic characteristics such as income (χ^2(6, N=2362) = 4.75, *p* = 0.58), educational attainment (χ^2(2, N=2782) = 1.92, *p* = 0.38), sex (χ^2(2, N=2782) = 1.48, *p* = 0.48), and age (by ANOVA; *F*(2, 2779) = 0.78, *p* = 0.46), indicative of successful randomization in assignment to the conditions. These and other covariates are further detailed in Table 2.

<<Table 1 about here>>

<<Table 2 about here>>

To test our hypothesis, we estimate an interval model for WTP using interval regression (Cameron and Huppert 1989), which is applicable to interval-censored data like those yielded by the payment scale. A linear relationship is modeled between WTP as a latent dependent variable and a set of covariates, with the cumulative standard normal density function being used for lower and upper interval limits and maximum likelihood estimation. Formally, in line with Cameron and Huppert (1989), we assume a linear relationship of the form $WTP_i = x_i'\beta + u_i$, where x_i' is a transposed vector of covariates, β is a vector of coefficients, and u_i is $NID(0, \sigma^2)$. Then, each pair of interval thresholds defined by the set of payment scale bids ($[c_n, c_{n+1}]$, n = 1, …, N-1) can be “standardized” to exploit the computational properties of probability calculus:

$$P\big(WTP_i \subseteq (c_{\mathrm{n-1}i}, c_{\mathrm{n}i})\big) = P\left(\frac{c_{\mathrm{n-1}i} - x_i'\beta}{\sigma} < z_i < \frac{c_{\mathrm{n}i} - x_i'\beta}{\sigma}\right), \qquad [1]$$

where z_i is the standard normal random variable. The joint probability density function for W independent observations of WTP can be thought of as a likelihood function for the unknown parameters β and σ (i.e., standard deviation of u). In this respect, the probability in [1] can be expressed as the difference between two standard normal cumulative densities, $\Phi(z_{\mathrm{u}i}) -$

$\Phi(z_{\mathrm{l}i})$, where $z_{\mathrm{u}i}$ and $z_{\mathrm{l}i}$ represent the fractions in [1] denoting the upper and lower limits, respectively. The log-likelihood function for the joint probability density is then given by

$$\log L = \sum_{i=1}^{W} \log[\Phi(z_{\mathrm{u}i}) - \Phi(z_{\mathrm{l}i})], \qquad [2]$$

where the unknown parameters β and σ are estimated using the maximum likelihood technique.

The vector of covariates is described in Table 2. The independent variable PAYMENTSCALE is of primary experimental interest, and standard sociodemographic covariates of income, age, education, and sex are included as control variables. Additional control variables include relevant respondent characteristics in terms of environmental attitudes (proxied by membership in environmental and/or recreational organizations), residential proximity to the sea, and previous experience with the environmental insults of the valuation scenarios. A final control variable (SMARTPHONE) is included owing to the potential confounding effects of smartphone usage on the Treatment conditions (Skeie et al. 2019). This concern arises because smartphone browser optimization resulted in only a few bids being clearly visible on the slider bar before interacting with the cursor.

Our WTP model is linear and additive in the set of included covariates. Therefore, observing the size and statistical significance of the coefficients of the categorical variable PAYMENTSCALE enables inference about the average effect on expected (latent) WTP of using each of the two Treatment payment scales, compared to the Control scale as a baseline. To test whether differences in the coefficients of the two Treatment scales are statistically significant, we employ standard post-estimation Wald tests.

For robustness purposes, we first re-estimate the interval regression WTP model after excluding speeders. We use the 5th and 10th percentile for the response time in minutes as thresholds for defining the fastest respondents (Sandorf, Aanesen, and Navrud 2016).

Subsequently, we re-estimate the WTP model using Tobit regression, with the selected payment scale bid as a conservative point estimate of WTP. The Tobit model relates the latent WTP to individual respondent characteristics, accounting for the global left-censoring of WTP data at zero. As outlined by Verbeek (2017), the model is defined as follows:

$$y_i^* = x_i'\beta + \varepsilon_i, \quad i = 1, 2, \ldots, N, \qquad [3]$$

$$y_i = y_i^* \quad \text{if } y_i^* > 0$$

$$= 0 \quad \text{if } y_i^* \leq 0$$

where y_i^* is the latent (unobserved) dependent variable, y_i is the censored (observed) dependent variable, x_i' is a vector of covariates, β is a vector of regression coefficients to be estimated, and ε_i is the error term, assumed to be $NID(0, \sigma^2)$ and independent of x_i.

We estimate the following Tobit model of WTP:

$$WTP^* = x_i'\beta + \varepsilon_i, \qquad [4]$$

with the same vector of covariates as for the interval regression. Maximum likelihood estimation is used to determine the vector of regression coefficients and the standard deviation of the error term. Regression coefficients will be presented as margins, denoting the marginal effects of covariates on the censored (observed) WTP, owing to the nonlinear relationship between raw Tobit coefficients and changes in the expected value of WTP conditional on observation.

Contingent valuation survey

We used an online contingent valuation survey conducted by Kantar TNS, an ISO-certified survey company, on its web panel in June 2022. This study replicates a 2015 study (Lindhjem et al. 2016) to update nonmarket value estimates for social benefit-cost analysis at the Norwegian Coastal Administration. The survey's methodology was also extensively described by Skeie et al. (2019), which forms the basis for the following description. The survey aimed to elicit the WTP for safety measures against oil spills due to ship accidents along the coastline of five regions in Norway and the resulting loss of ecosystem services. In this study, we focus on data collected from the "eastern" region, including four valuation scenarios in the Oslofjord area.

The survey informed respondents that, in the absence of new safety measures, the increasing coastal traffic in Norway could likely result in an oil spill within the next few years. Visual aids, such as maps, were used to demonstrate four potential levels of oil dispersion. These levels were determined using oil spill dispersion modeling (Jødestøl et al. 2001) and expert inputs. The environmental damages and losses in ecosystem services associated with each oil spill scenario were presented in a table. This table outlined the key impacts on seabirds, seals, general marine life, and the coastline in the Oslofjord area, as illustrated in Figure 2. To reduce cognitive load, we maintained consistent color codes between the map dispersion and damage levels in the table. While the term "ecosystem services" was not explicitly mentioned, the focus was evidently on recreational services, including the consumption of healthy, self-caught seafood, and nonuse values related to coastal ecosystem protection, including biodiversity and specific species.

<<Fig 2 about here>>

Respondents were invited to indicate their WTP to avoid impacts from the least to the most severe damage levels. This process followed an advanced disclosure procedure (Bateman et al. 2004) and included a "cheap talk" script to remind them of their budget constraints. Each scenario was displayed on a single screen, showcasing the ecosystem service damage table with one of the four damage columns highlighted (and the other three dimmed), along with the corresponding oil dispersion map. The payment vehicle was a one-time household tax allocated exclusively for measures guaranteed to prevent such damages. Before initiating the first data collection round, the survey's design, including the elicitation format, were extensively tested over several years through pilot studies, focus groups, and personal interviews with survey respondents (Navrud, Lindhjem, and Magnussen 2017).

Internal and external validity

A prerequisite for internal validity in contingent valuation studies is the incentive compatibility of the payment elicitation format (Carson and Groves 2007). In keeping with the mechanism design theory, Carson and Groves argued that formats offering more than two alternatives generally lack incentive compatibility. However, Vossler and Holladay (2018) have shown how to achieve theoretical incentive compatibility in popular payment elicitation formats, including payment cards, by four conditions First, respondents must care about the outcome. Second, the authority must be able to enforce payments by the voters (payment consequentiality). Third, the elicitation must resemble a yes or no vote for a single project. Fourth, the probability of project implementation must be weakly monotonically increasing with the proportion of yes votes (policy consequentiality). Assuming that the respondent's true valuation of the project is *v*, compliance with the conditions will incentivize selecting the payment card bid representing the lower bound of the interval containing *v*, denoted by [c_v, c_{v+1}].

In the current case, respondents have demonstrably cared about the outcome of our contingent valuation survey, by responding to a question on perceived loss of life quality associated with the ecosystem service damages prior to WTP elicitation (Lindhjem et al. 2016). The survey in the present research also used a tax as the payment vehicle to ensure payment consequentiality. Vossler and Holladay (2018) state that circling the highest amount one is willing to pay in a payment card can be interpreted as a series of yes and no votes, as outlined in Section 2. Our payment scale may better facilitate the cognitive process of yes and no voting by its horizontal presentation of bids, compared to the columns and rows of the conventional payment card. Compliance with the fourth condition depends on (respondents' belief about) the nature of vote interpretation if the actual project cost resides in the interval implied by the selected bid and the next bid, [c_v, c_{v+1}]. While our survey did not include explicit information on this aspect, we hold that incentive compatibility is reasonably maintained overall in our preference elicitation.

Regarding the internal validity of our study design, interval regression is the preferred estimation method for modelling payment card WTP data (Boyle 2003). Additionally, the Tobit regression model provides a useful robustness check of our results by accommodating the left-censoring at zero in the observed WTP responses. We suggest that the overall consequentiality of the contingent valuation survey, combined with the split-sample experimental design and our chosen empirical approach, collectively provide adequate internal validity for drawing causal inferences from our findings.

The current study exhibits strong external validity in several aspects. The visual representation of the impacts on ecosystem services, derived by oil spill dispersion modelling, closely resembles those in standard contingent valuation surveys, ensuring a high degree of experimental realism (Morales, Amir, and Lee 2017). Regarding ecological validity,[6] our study aligns with the framework of a natural field experiment (Harrison and List 2004). First, it examines the potential impacts of payment scale design on the stated WTP in contingent valuation surveys by subjecting respondents to experimental conditions within the actual environment we aim to generalize our results to. Second, the respondents are not aware of the specific experimental conditions related to the elicitation format. Regarding sampling validity,[7] Kantar TNS sustains its web panel at a target size of 40,000 members through probabilistic recruitment in nationally representative surveys. This panel employs stratification to ensure the sample's representativeness of the population of interest. The response rate for the current data collection was 34%, which is consistent with our expectations for current web panel response rates.

Consequently, the ecological and sampling validity of this study should support broad generalizations of the effects of bid design variation on stated preferences to both domestic and international contingent valuation web survey respondents encountering a payment scale elicitation format. This generalizability holds to the extent that observed effects reflect fundamental aspects of the human decision-making process. We suggest that the results, at least to some extent, also generalize to the conventional payment card format.

4. Results

Descriptive statistics of WTP, with the selected payment scale bid coded as a point estimate of WTP, are given in Table 3. The mean WTP increases across the payment scale conditions as expected for all damage levels but “very large,” where that of the Treatment 2 condition is slightly lower than that of the Treatment 1 condition. The median WTP, which is the median selected bid, exhibits the same pattern, although both the Control and Treatment 2 conditions exceed the Treatment 1 condition for “very large” damage. Finally, the largest stated WTP exceeds the highest payment scale bid for all three conditions for “medium,” “large,” and “very large” damage (elicited through an open-ended follow-up question when selecting “more than [NOK10,000/NOK12,000]”). The maximum stated WTP associated with the Treatment 1 scale (NOK100,000) greatly exceeds those associated with the Control (NOK40,000) and Treatment 2 (NOK30,000) scales. However, we believe this is coincidental rather than being scale-induced, as the highest bid is equal between the Treatment conditions.

<<Table 3 about here>>

Turning to the interval model for WTP, our hypothesis is supported by the fact that we observed significant effects on mean WTP for each of the Treatment conditions, compared to the Control condition as a baseline, as shown in Table 4. For the first WTP elicitation (i.e., "small" ecosystem service damage), the increase in the bid amounts on the first part of the payment scale in Treatment 1 yields a significantly higher mean ($\beta_{\text{Treatment 1}}$ = 189.79, p = 0.008) compared to that of the Control payment scale. The additional increase in the bid amounts in the Treatment 2 condition yields an even more pronounced increase in mean WTP compared to that of the Control condition ($\beta_{\text{Treatment 2}}$ = 326.78, p < 0.001). The additional increase in mean WTP for the Treatment 2 condition is significant compared to that of the Treatment 1 condition, as found by a post-elicitation Wald test: χ^2(1, N = 1932) = 3.91, p = 0.048.

<<Table 4 about here>>

The differences in mean WTP between the Control and the two Treatment conditions are amplified for the scope programs, both for "medium" ($\beta_{\text{Treatment 1}}$ = 344.91, p < 0.001; $\beta_{\text{Treatment 2}}$ = 432.60, p < 0.001) and "large" ($\beta_{\text{Treatment 1}}$ = 561.75, p < 0.001; $\beta_{\text{Treatment 2}}$ = 627.31, p < 0.001) ecosystem service damages. However, the difference in the coefficients for the Treatment 1 and Treatment 2 conditions are no longer significant for both elicitations ("medium": χ^2(1, N = 1911) = 0.91, p = 0.34; "large": χ^2(1, N = 1903) = 0.27, p = 0.60). The fourth elicitation ("very large" damage) yields the highest elicitation effect in our study when comparing the Treatment 1 and Control conditions ($\beta_{\text{Treatment 1}}$ = 668.53, p < 0.001), whereas Treatment 2 shows a somewhat lower, but still highly significant, effect on mean WTP compared to the Control condition for this scope program ($\beta_{\text{Treatment 2}}$ = 565.96, p = 0.001). Meanwhile, the difference in coefficients between the two Treatment conditions is not significant (χ^2(1, N = 1864) = 0.42, p = 0.52).

In terms of effect sizes, the largest coefficients reported above (for the "very large" ecosystem service damage) indicate that expected WTP increase by NOK669 and NOK566 when elicited by the Treatment 1 and Treatment 2 payment scales, respectively, compared to the Control one. While the former constitutes the largest nominal elicitation effect, the largest relative effect arises for the "small" damage level. Here, the increase in the predicted mean WTP from the interval regression model is 43% when comparing the Control condition (NOK752) to the Treatment 2 condition (NOK1079). Kernel density plots comparing the estimated WTP distributions across the experimental conditions show clear distributional shifts in the same direction. The plots are based on the predicted WTP from the interval regression model and are provided in Appendix Figure 1-4.

The elicitation effects remain robust when excluding speeders. Neither the structure nor the size of significant coefficients changed in substantial ways for any of the two exclusion stages (respondents below the 5^{th} and 10^{th} percentile of response time, respectively), as shown in Appendix Table A1 and Appendix Table A2.

The results are also robust to re-estimation of the WTP model with Tobit regression when using the selected bid as a point estimate of the WTP, as shown in Table 5. The coefficients for both Treatment conditions are somewhat reduced for all four elicitations but remain significant at the 1% level in all cases but two ("small": $\beta_{\text{Treatment 1}}$ = 92.18, p = 0.064; "very large": $\beta_{\text{Treatment 2}}$ = 340.20, p = 0.019). Regarding response validity, the coefficients on INCOMELEVEL21 are positive and largely significant at the 1% level, as expected, with both estimation methods. Additionally, the positive and significant coefficients for our proxy of environmental attitudes, indicated by membership in environmental or recreational organizations (MEMBER) as well as previous experience with environmental damages due to oil spills (PREVIOUSEXP), also perform as expected for both estimations. Finally, no evidence was found of an effect of smartphone usage on the stated WTP. Collectively, the Tobit regression results corroborate the findings from the interval regression, and construct validity is indicated with both estimation methods.

<<Table 5 about here>>

5. Discussion

Main findings

In a split-sample experiment through a contingent valuation survey to determine respondents' WTP for avoiding ecosystem service damages due to coastal oil spills, we found clear evidence of anchoring effects on stated preferences in the payment card elicitation format. By interval regression, we found that higher initial bids significantly increased the mean WTP across three payment card versions, and this was persistent through a sequence of four elicitations of the WTP. Kernel density plots further showed clear shifts in WTP distributions in the same direction. The findings are robust to re-estimation after excluding speeders and re-estimation with Tobit regression.

Following the anchoring-and-adjustment heuristic (Tversky and Kahneman 1974; Epley and Gilovich 2006), we suggest that respondents anchor on one or several bids in the first part of the payment scale and adjust from there to state their maximum WTP. Consequently, anchoring on lower bids in the Control condition coupled with insufficient adjustment might account for the lower observed mean WTP in comparison to the two Treatment conditions. The explanatory weight on the first part of the payment scale is substantiated by the fact that both of the Treatment conditions had lower endpoint bids and Treatment 1 had a lower midpoint bid, respectively, compared to the Control condition, while still producing higher mean WTP and right-ward shifts in the WTP distribution.

Methodological implications

The current study demonstrates that bid design variations within the payment card elicitation format significantly influence the stated preferences. This indicates the susceptibility of preferences to psychological cues. While the phenomenon of anchoring on a single bid is well-established in dichotomous choice formats and bidding games (Boyle, Bishop, and Welsh 1985; Herriges and Shogren 1996; Green et al. 1998; Frew, Wolstenholme, and Whynes 2004), the potential for an anchoring process within the payment card format is less apparent owing to the variety of potential anchors, as noted by Boyle (2003). Nonetheless, given that some extent of anchoring on the set of bids, indicatively the first few, may be inevitable, mitigation of the influence of the practitioners' own bid design choices on environmental value estimates warrants more attention.

Practical implications

We initially establish the economic significance of our findings before discussing their implications for contingent valuation practitioners. Regarding economic significance, Funder and Ozer (2019) argue that even minor individual-level effects can accumulate over time or among individuals, leading to substantial aggregate effects, provided the results of an intervention are robust, replicable, and scalable at a population level (see also List 2022). In social benefit-cost analysis, individual value estimates from stated preference studies are commonly aggregated across the relevant affected population to determine welfare changes resulting from program implementation. The extent to which welfare estimates are sensitive to variations in the mean WTP depends on the affected population's size, which is generally substantial when significant nonuse values are present. Consequently, a slight difference in mean WTP can become large when multiplied across millions of citizens.

The economic significance of the results can be demonstrated by a simple calculation. The respective affected populations for the four ecosystem service damage scenarios in the contingent valuation survey are defined in the social benefit-cost analysis model at the Norwegian Coastal Administration. This enables analyses of sensitivity in aggregate welfare estimates to the differences in mean WTP evident in the current study. For the "small" ES damage level, the increase in the predicted mean WTP from the Control condition (NOK752) to the Treatment 2 condition (NOK1079) is 43%. This increase translates into a MNOK 234 difference when aggregated over the relevant affected population.[8] This difference is approximately equivalent to $20 million USD. For “very large” ecosystem service damage, a 24% increase in the predicted mean WTP from the Control condition (NOK 2328) to the Treatment 2 condition (NOK 2894) results in a MNOK 829 difference in welfare, which is approximately equivalent to $73 million USD.[9] To provide context, this latter difference is more than 50% larger than the entire welfare estimate for “small” ecosystem service damage using the predicted mean WTP for the Control condition.

The implications of context dependence in stated preferences for aggregate welfare measures raises practical design questions. This study focuses on the payment card format, where the existence of a definitive, normative bid design is unclear. This uncertainty stems from the inherent invisibility of true preferences and the potential context dependence of stated preferences. Although it is impossible to determine if stated preferences diverge from true, unobserved preferences in any specific study design, employing a split-sample study with bid design variation can at least minimize the influence of the practitioner’s selection of a singular design on the value estimates.

Decisions on split-sample design are influenced by several interconnected factors. Primarily, the available budget sets a limit on the feasible total sample size. Additionally, decisions on the number of conditions to include in a split-sample design depends on the desired precision in WTP estimates, given the total sample size. Even with ample funding, the size of the relevant sampling frame also restricts the total feasible sample size. This limitation is evident in cases like sampling from an online panel, where the response rate also plays an important role. When facing budgetary or sampling frame constraints, resulting in a smaller feasible total sample size, the practitioner might need to balance accommodating context-dependent preferences against the precision of WTP estimation. Including multiple experimental conditions with bid design variation can increase the variance of the mean WTP estimate in each condition for a given, feasible total sample size. Therefore, final decisions on study design should involve a careful consideration of both these aspects.

Limitations and future research

One limitation of our study is the different payment scale endpoints in the Control condition (NOK12,000) compared to the two Treatment conditions (NOK10,000), which may confound anchoring effects with range effects. As outlined in Section 2, Rowe, Schulze, and Breffle (1996) found that varying endpoints in the payment card elicitation format did not affect the observed WTP distribution until the endpoint was low enough to truncate the distribution of the true WTP. Based on the observed share of maximum bid selection in our study[10], we believe each of the two endpoints are high enough to not significantly truncate the distribution of the true WTP. Moreover, even though later studies have found evidence of range effects (Covey, Loomes, and Bateman 2007; Soeteman, van Exel, and Bobinac 2017), the bids on the first part of the payment cards in these studies are higher when the range is higher, which may well confound range effects with anchoring effects. All in all, we believe our difference in endpoints does not invalidate our results, especially not for the first part of the payment card, which was the focus of the current research.

Another potential limitation is the difference in elicitation between the survey of the Control condition and that of the two Treatment conditions, asking for household and individual WTP, respectively. Ex ante, household WTP should weakly exceed individual WTP for identical environmental programs. Our results show that household WTP in the Control condition is consistently lower than individual WTP in any of the Treatment conditions. The observed experimental differences in our study may therefore, if anything, be somewhat attenuated.

In terms of fruitful future research avenues, extending our study to explore whether the progressivity of the set of payment card bids also influences stated preferences—in addition to anchoring on initial bids—will further advance our understanding of elicitation effects in the payment card format. A concentration of granulated bids in a certain part of the payment card may itself serve as an anchor as it conveys the researcher's expectation of the likely region in which most people's WTP resides. Second, an exploration of how respondents interact with the payment card would be beneficial for the mitigation of elicitation effects. For instance, by using eye-tracking technology with online surveys, one could assess what parts of the payment card attracts most attention from respondents before deciding on a statement of maximum WTP. Alternatively, one could include an item battery in a contingent valuation survey exploring whether and which features of the payment card are formative for stated preferences.

Beyond the payment card elicitation format, we also recommend increased research into the effects of design variation in describing ecosystem service attribute levels in contingent valuation surveys. Ahi, Aanesen, and Kipperberg (2023) observed that simple attribute translations[11] in a discrete choice experiment significantly impacted stated preferences, suggesting that the contingent valuation method might similarly be susceptible to the effects of subtle design variation in features other than the payment elicitation format. A research agenda focused on systematically exploring how stated preferences, and in turn, aggregate welfare estimates, are influenced by subtle design variations at the discretion of the individual researcher and practitioner would further strengthen the stated preference methods as tools for environmental valuation.

6. Conclusion

This study investigates the effects of payment card bid design variation on stated preferences using a split-sample experiment in a contingent valuation survey. Three experimental conditions with different bid designs produced significantly different mean WTP estimates. This difference persisted through a sequence of four elicitations of WTP for avoiding coastal environmental damages of increasing size. The main methodological implication of the results is that the stated preferences elicited by the payment card are significantly influenced by the bid design. The economic significance of the results is highlighted through a sensitivity analysis of welfare estimates by using the social benefit-cost analysis model of a national government agency. Our findings suggest that contingent valuation practitioners using the payment card format should consider accommodating context-dependence in stated preferences by incorporating split-sample experiments with bid design variation in contingent valuation studies. This approach minimizes the impact of a practitioner's choice of a single bid design on value estimates. Our results underscore the need for further investigation into the effects of subtle design variation on the stated preferences in contingent valuation surveys.

7. Acknowledgements

We acknowledge funding from the Research Council of Norway (grant numbers 327210 and 319917). We also gratefully acknowledge the helpful comments and advice from attendees at the IAREP Conference 2022, the Behavioral Environmental Economics Workshop 2022 hosted by the Swedish University of Agricultural Sciences, and the 44th Annual Meeting of the Norwegian Association of Economists. Finally, we thank Gorm Kipperberg for initial discussions on the role of design variations in stated preference research.

References

Ahi, Jülide Ceren, Margrethe Aanesen, and Gorm Kipperberg. 2023. 'Testing the Sensitivity of Stated Environmental Preferences to Variations in Choice Architecture'. *Ecological Economics* 205 (March). https://doi.org/10.1016/j.ecolecon.2022.107680.

Akinc, Deniz, Deborah J. Street, and Martina Vandebroek. 2024. 'Varying Choice Set Sizes in Discrete Choice Experiments'. *Journal of Choice Modelling* 52 (September). https://doi.org/10.1016/j.jocm.2024.100493.

Alberini, Anna, Kevin Boyle, and Michael Welsh. 2003. 'Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty'. *Journal of Environmental Economics and Management* 45:40–62. https://doi.org/10.1016/S0095-0696(02)00010-4.

Alevy, Jonathan E., John A. List, and Wiktor L. Adamowicz. 2011. 'How Can Behavioral Economics Inform Nonmarket Valuation? An Example from the Preference Reversal Literature'. *Land Economics* 87 (3): 365–81. https://doi.org/10.3368/le.87.3.365.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. '"Coherent Arbitrariness": Stable Demand Curves without Stable Preferences'. *The Quarterly Journal of Economics* 118 (1): 73–105. https://doi.org/10.1162/00335530360535153.

Bateman, Ian J., Matthew Cole, Philip Cooper, Stavros Georgiou, David Hadley, and Gregory L. Poe. 2004. 'On Visible Choice Sets and Scope Sensitivity'. *Journal of Environmental Economics and Management* 47 (1): 71–93. https://doi.org/10.1016/S0095-0696(03)00057-3.

Bateman, Ian J, Ian H Langford, R Kerry Turner, Ken G Willis, and Guy D Garrod. 1995. 'Elicitation and Truncation Effects in Contingent Valuation Studies'. *Ecological Economics* 12:161–79. https://doi.org/10.1016/0921-8009(94)00044-V.

Bishop, Richard C., Kevin J. Boyle, Richard T. Carson, David Chapman, W. Michael Hanemann, Barbara Kanninen, Raymond J. Kopp, Jon A. Krosnick, John List, Norman Meade, Robert Patterson, Stanley Presser, V. Kerry Smith, Roger Tourangeau, Michael Welsh, Jeffrey M. Wooldridge, Matthew DeBell, Colleen Donovan, Matthew Konopka, Nora Scherer. 2017. 'Putting a Value on Injuries to Natural Assets: The BP Oil Spill'. *Science* 356 (6335): 253–54. https://doi.org/10.1126/science.aam8124.

Blaine, Thomas W., Frank R. Lichtkoppler, Keith R. Jones, and Randall H. Zondag. 2005. 'An Assessment of Household Willingness to Pay for Curbside Recycling: A Comparison of Payment Card and Referendum Approaches'. *Journal of Environmental Management* 76 (1): 15–22. https://doi.org/10.1016/j.jenvman.2005.01.004.

Börger, Tobias, Tenaw G. Abate, Margrethe Aanesen, and Ewa Zawojska. 2021. 'Payment and Policy Consequentiality in Dichotomous Choice Contingent

Valuation: Experimental Design Effects on Self-Reported Perceptions'. *Land Economics* 97 (2): 407–24. https://doi.org/10.3368/le.97.2.407.

Boyle, Kevin J. 2003. 'Contingent Valuation in Practice'. In *A Primer on Nonmarket Valuation*, edited by Patricia A Champ and Thomas C Brown, 111–69. New York: Springer Science+Business Media, LLC.

Boyle, Kevin J, Richard C Bishop, and Michael P Welsh. 1985. 'Starting Point Bias in Contingent Valuation Bidding Games'. *Land Economics* 61 (2): 188–94. https://www.jstor.org/stable/3145811.

Boyle, Kevin J, Hugh F Macdonald, Hsiang-Tai Cheng, and Daniel W Mccollum. 1998. 'Bid Design and Yea Saying in Single-Bounded, Dichotomous-Choice Questions'. *Land Economics* 74 (1): 49–64. https://www.jstor.org/stable/3147212.

Cameron, Trudy Ann, and Daniel D Huppert. 1989. 'OLS versus ML Estimation of Non-Market Resource Values with Payment Card Interval Data'. *Journal of Environmental Economics and Management* 17:230–46. https://doi.org/10.1016/0095-0696(89)90018-1.

Carson, Richard T, Nicholas E Flores, and Norman F Meade. 2001. 'Contingent Valuation: Controversies and Evidence'. *Environmental and Resource Economics* 19:173–210. https://doi.org/10.1023/A:1011128332243.

Carson, R T, R C Mitchell, W M Hanemann, R J Kopp, S Presser, and P A Ruud. 1992. 'A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill'. Munich. https://mpra.ub.uni-muenchen.de/6984/.

Carson, Richard T, Robert C Mitchell, Michael Hanemann, Raymond J Kopp, Stanley Presser, and Paul A Ruud. 2003. 'Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill'. *Environmental and Resource Economics* 25:257–86. https://doi.org/10.1023/A:1024486702104.

Carson, Richard T, and Theodore Groves. 2007. 'Incentive and Informational Properties of Preference'. *Environmental and Resource Economics* 37:181–210. https://doi.org/10.1007/s10640-007-9124-5.

Carson, Richard T, Leanne Wilks, and David Imber. 1994. 'Valuing the Preservation of Australia's Kakadu Conservation Zone'. *Oxford Economic Papers* 46:727–49. https://www.jstor.org/stable/2663496.

Carthy, Trevor, Susan Chilton, Judith Covey, Lorraine Hopkins, Michael Jones-Lee, Graham Loomes, Nick Pidgeon, and Anne Spencer. 1999. 'On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 2-The CV/SG "Chained" Approach'. *Journal of Risk and Uncertainty* 17 (3): 187–213. https://doi.org/10.1023/A:1007782800868.

Champ, Patricia A, and Richard C Bishop. 2006. 'Is Willingness to Pay for a Public Good Sensitive to the Elicitation Format?' *Land Economics* 82 (2): 162–73. https://doi.org/10.3368/le.82.2.162.

Champonnois, Victor, Olivier Chanel, and Khaled Makhloufi. 2018. ‘Reducing the Anchoring Bias in Multiple Question CV Surveys’. *Journal of Choice Modelling* 28 (September):1–9. https://doi.org/10.1016/J.JOCM.2018.04.005.

Chanel, Olivier, Khaled Makhloufi, and Mohammad Abu-Zaineh. 2017. ‘Can a Circular Payment Card Format Effectively Elicit Preferences? Evidence From a Survey on a Mandatory Health Insurance Scheme in Tunisia’. *Applied Health Economics and Health Policy* 15 (3): 385–98. https://doi.org/10.1007/s40258-016-0287-5.

Chestnut, Lauraine G., Robert D. Rowe, and William S. Breffle. 2012. ‘Economic Valuation of Mortality-Risk Reduction: Stated Preference Estimates from the United States and Canada’. *Contemporary Economic Policy* 30 (3): 399–416. https://doi.org/10.1111/j.1465-7287.2011.00269.x.

Clark, Jeremy, and Lana Friesen. 2008. ‘The Causes of Order Effects in Contingent Valuation Surveys: An Experimental Investigation’. *Journal of Environmental Economics and Management* 56 (2): 195–206. https://doi.org/10.1016/j.jeem.2007.12.005.

Cooper, Joseph, and John Loomis. 1992. ‘Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models’. *Land Economics* 68 (2): 211–24. https://www.jstor.org/stable/3146775.

Covey, Judith, Graham Loomes, and Ian J. Bateman. 2007. ‘Valuing Risk Reductions: Testing for Range Biases in Payment Card and Random Card Sorting Methods’. *Journal of Environmental Planning and Management* 50 (4): 467–82. https://doi.org/10.1080/09640560701401986.

Dietz, Thomas, Paul C Stern, and Amy Dan. 2009. ‘How Deliberation Affects Stated Willingness to Pay for Mitigation of Carbon Dioxide Emissions: An Experiment’. *Land Economics* 85 (2): 329–47. https://muse.jhu.edu/article/466902/summaryhttps://muse.jhu.edu/related_content?type=article&id=466902.

Dubourg, W R, M W Jones-Lee, and Graham Loomes. 1997. ‘Imprecise Preferences and Survey Design in Contingent Valuation’. *Economica* 64:681–702. https://doi.org/10.1111/1468-0335.00106.

Epley, Nicholas, and Thomas Gilovich. 2006. ‘The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient’. *Psychological Science* 17 (4): 311–18. https://doi.org/10.1111/j.1467-9280.2006.01704.x.

Faccioli, Michela, and Klaus Glenk. 2022. ‘More in Good Condition or Less in Bad Condition? Valence-Based Framing Effects in Environmental Valuation’. *Land Economics* 98 (2): 314–36. https://doi.org/10.3368/le.98.2.051920-0067R1.

Faul, Franz, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. ‘G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences’. *Behavior Research Methods* 39 (2): 175–91. https://doi.org/10.3758/BF03193146.

Fischhoff, Baruch. 1991. ‘Value Elicitation. Is There Anything in There?’ *American Psychologist* 46 (8): 835–47. https://doi.org/10.1037/0003-066X.46.8.835.

Frew, Emma J., David K. Whynes, and Jane L. Wolstenholme. 2003. ‘Eliciting Willingness to Pay: Comparing Closed-Ended with Open-Ended and Payment Scale Formats’. *Medical Decision Making* 23 (2): 150–59. https://doi.org/10.1177/0272989X03251245.

Frew, Emma J., Jane L. Wolstenholme, and David K. Whynes. 2004. ‘Comparing Willingness-to-Pay: Bidding Game Format versus Open-Ended and Payment Scale Formats’. *Health Policy* 68 (3): 289–98. https://doi.org/10.1016/j.healthpol.2003.10.003.

Frey, Ulrich J., and Frauke Pirscher. 2018. ‘Willingness to Pay and Moral Stance: The Case of Farm Animal Welfare in Germany’. *PLoS ONE* 13 (8). https://doi.org/10.1371/journal.pone.0202193.

Funder, David C., and Daniel J. Ozer. 2019. ‘Evaluating Effect Size in Psychological Research: Sense and Nonsense’. *Advances in Methods and Practices in Psychological Science* 2 (2): 156–68. https://doi.org/10.1177/2515245919847202.

Grant, Laura, and Christian Langpap. 2020. ‘Willingness to Contribute as a Component of the Social Cost of Water Pollution’. *Land Economics* 96 (4): 573–88. https://doi.org/10.3368/wple.96.4.573.

Green, Donald, Karen E Jacowitz, Daniel Kahneman, and Daniel Mcfadden. 1998. ‘Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods’. *Resource and Energy Economics* 20:85–116. https://doi.org/10.1016/S0928-7655(97)00031-6.

Gregory, Robin, Sarah Lichtenstein, and Paul Slovic. 1993. ‘Valuing Environmental Resources: A Constructive Approach’. *Journal of Risk and Uncertainty* 7:177–97. https://doi.org/10.1007/BF01065813.

Harrison, Glenn W, and John A List. 2004. ‘Field Experiments’. *Journal of Economic Literature* 42 (4): 1009–55. https://doi.org/10.1257/0022051043004577.

Herriges, Joseph A, and Jason F Shogren. 1996. ‘Starting Point Bias in Dichotomous Choice Valuation with Follow-Up Questioning’. *Journal of Environmental Economics and Management* 30:112–31. https://doi.org/10.1006/jeem.1996.0008.

Holleman, Gijs A., Ignace T.C. Hooge, Chantal Kemner, and Roy S. Hessels. 2020. ‘The “Real-World Approach” and Its Problems: A Critique of the Term Ecological Validity’. *Frontiers in Psychology* 11 (April):1–12. https://doi.org/10.3389/fpsyg.2020.00721.

Holmes, Thomas P., and Randall A. Kramer. 1995. ‘An Independent Sample Test of Yea-Saying and Starting Point Bias in Dichotomous Choice Contingent Valuation’. *Journal of Environmental Economics and Management* 29:121–32. https://doi.org/10.1006/jeem.1995.1035.

Jensen, Anne Kejser, Robert J. Johnston, and Søren B. Olsen. 2019. 'Does One Size Really Fit All? Ecological Endpoint Heterogeneity in Stated Preference Welfare Analysis'. *Land Economics* 95 (3): 307–32. https://doi.org/10.3368/le.95.3.307.

Jødestøl, Kjell Andreas, Bjørge Fredheim, Espen Edward Hoell, Sami Wakili, Jan Erik Vinnem, Arne Uno Myhrvold, Jon Rytter Hasle, and Gina Ytterborg. 2001. 'Achieving an Industry Standard in the Assessment of Environmental Risk: Oil Spill Risk Management and the Mira Method'. In *International Oil Spill Conference Proceedings*, 155–65. http://meridian.allenpress.com/iosc/article-pdf/2001/1/155/1747507/2169-3358-2001-1-155.pdf.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennet, Roy Brouwer, Trudy A. Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau, Christian A. Vossler. 2017. 'Contemporary Guidance for SP Studies'. *Journal of the Association of Environmental and Resource Economists* 4 (2): 319–405. https://doi.org/10.1086/691697.

Kahneman, Daniel, Ilana Ritov, Karen E Jacowitz, and Paul Grant. 1993. 'Stated Willingness to Pay for Public Goods: A Psychological Perspective'. *Psychological Science* 4 (5): 310–15. https://doi.org/10.1111/j.1467-9280.1993.tb00570.x.

Kanninen, Barbara J. 1995. 'Bias in Discrete Response Contingent Valuation'. *Journal of Environmental Economics and Management* 28:114–25. https://doi.org/10.1006/jeem.1995.1008.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. 'From Exxon to BP: Has Some Number Become Better than No Number?' *Journal of Economic Perspectives* 26 (4): 3–26. https://doi.org/10.1257/jep.26.4.3.

Li, Tongzhe, Jacob R. Fooks, Kent D. Messer, and Paul J. Ferraro. 2021. 'A Field Experiment to Estimate the Effects of Anchoring and Framing on Residents' Willingness to Purchase Water Runoff Management Technologies'. *Resource and Energy Economics* 63 (February):1–10. https://doi.org/10.1016/J.RESENEECO.2019.07.001.

Lichtenstein, Sarah, and Paul Slovic. 2009. 'The Construction of Preference: An Overview'. In *The Construction of Preference*, edited by Sarah Lichtenstein and Paul Slovic, 1–40. Cambridge University Press. https://doi.org/10.1017/cbo9780511618031.002.

Lindhjem, Henrik, Kristin Magnussen, Ståle Navrud, Sofie Skjeflo, and Odd Willy Brude. 2016. 'Verdsetting Av Miljørelatert Velferdstap Ved Oljeutslipp Fra Skip: Kalkulasjonspriser for Samfunnsøkonomiske Analyser'. Oslo.

Lindhjem, Henrik, and Ståle Navrud. 2009. 'Asking for Individual or Household Willingness to Pay for Environmental Goods? : Implication for Aggregate Welfare Measures'. *Environmental and Resource Economics* 43 (1): 11–29. https://doi.org/10.1007/s10640-009-9261-0.

Lindhjem, Henrik, and Ståle Navrud. 2011. 'Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes'. *International Review of*

Environmental and Resource Economics 5 (4). https://doi.org/10.1561/101.00000045.

List, John A. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. New York: Penguin Random House.

Lusk, Jayson L, and F Bailey Norwood. 2009. 'An Inferred Valuation Method'. *Land Economics* 85 (3): 500–514. https://doi.org/10.3368/le.85.3.500.

Luther, Zachary R., Scott M. Swinton, and Braeden Van Deynze. 2022. 'Potential Supply of Midwest Cropland for Conversion to In-Field Prairie Strips'. *Land Economics* 98 (2): 274–91. https://doi.org/10.3368/le.98.2.082020-0129R1.

Mahieu, Pierre-Alexandre, Romain Crastes dit Sourd, Pascal Gastineau, Benoît Chèze, and Andrea Taylor. 2024. 'How Does the Cost Vector Affect Payment Consequentiality in a Binary Choice Contingent Valuation Survey?' *Land Economics* 100 (2): 245–55. https://doi.org/10.3368/le.100.2.111821-0135r1.

Maineri, Angelica M., Ivano Bison, and Ruud Luijkx. 2021. 'Slider Bars in Multi-Device Web Surveys'. *Social Science Computer Review* 39 (4): 573–91. https://doi.org/10.1177/0894439319879132.

McCloskey, Deirdre N, and Stephen T Ziliak. 1996. 'The Standard Error of Regressions'. *Journal of Economic Literature* 34 (1): 97–114. https://www.jstor.org/stable/2729411.

McEwan, Bree. 2020. 'Sampling and Validity'. *Annals of the International Communication Association* 44 (3): 235–47. https://doi.org/10.1080/23808985.2020.1792793.

Meyerhoff, Jürgen, Malte Oehlmann, and Priska Weller. 2015. 'The Influence of Design Dimensions on Stated Choices in an Environmental Context'. *Environmental and Resource Economics* 61 (3): 385–407. https://doi.org/10.1007/s10640-014-9797-5.

Mitchell, Robert C, and Richard T Carson. 1981. 'An Experiment In Determining Willingness To Pay For National Water Quality Improvements'. Washington, D.C.

Mitchell, Robert C., and Richard T. Carson. 1989. 'Using Surveys to Value Public Goods: The Contingent Valuation Method'. In . Washington, D.C.: Resources for the Future.

Morales, Andrea C., On Amir, and Leonard Lee. 2017. 'Keeping It Real in Experimental Research-Understanding When, Where, and How to Enhance Realism and Measure Consumer Behavior'. *Journal of Consumer Research* 44 (2): 465–76. https://doi.org/10.1093/jcr/ucx048.

Navrud, Ståle, Henrik Lindhjem, and Kristin Magnussen. 2017. 'Valuing Marine Ecosystem Services Loss from Oil Spills for Use in Cost-Benefit Analysis of Preventive Measures'. In *Handbook on the Economics and Management for Sustainable Oceans*, 124–37. Cheltenham: Edward Elgar Publishing.

Ndambiri, Hilary, Roy Brouwer, and Eric Mungatana. 2016. 'Comparing Welfare Estimates across Stated Preference and Uncertainty Elicitation Formats for Air

Quality Improvements in Nairobi, Kenya'. *Environment and Development Economics* 21 (5): 649–68. https://doi.org/10.1017/S1355770X15000455.

Ndambiri, Hilary, Eric Mungatana, and Roy Brouwer. 2017. 'Scope Effects of Respondent Uncertainty in Contingent Valuation: Evidence from Motorized Emission Reductions in the City of Nairobi, Kenya'. *Journal of Environmental Planning and Management* 60 (1): 22–46. https://doi.org/10.1080/09640568.2016.1140024.

Payne, John W, James R Bettman, and Eric J Johnson. 1992. 'Behavioral Decision Research: A Constructive Processing Perspective'. *Annual Review of Psychology* 43:87–131. https://doi.org/10.1146/annurev.ps.43.020192.000511.

Powe, N. A., and I. J. Bateman. 2003. 'Ordering Effects in Nested "Top-Down" and "Bottom-Up" Contingent Valuation Designs'. *Ecological Economics* 45 (2): 255–70. https://doi.org/10.1016/S0921-8009(03)00084-3.

Ready, Richard C, Ståle Navrud, and W Richard Dubourg. 2001. 'How Do Respondents with Uncertain Willingness to Pay Answer Contingent Valuation Questions?' *Land Economics* 77 (3): 315–26. https://doi.org/10.2307/3147126.

Reaves, Dixie Watts, Randall A Kramer, and Thomas P Holmes. 1999. 'Does Question Format Matter? Valuing an Endangered Species'. *Environmental and Resource Economics* 14:365–83. https://doi.org/10.1023/A:1008320621720.

Roach, Brian, Kevin J. Boyle, and Michael Welsh. 2002. 'Testing Bid Design Effects in Multiple-Bounded, Contingent-Valuation Questions'. *Land Economics* 78 (1): 121–31. https://doi.org/10.2307/3146927.

Rowe, Robert D, William D Schulze, and William S Breffle. 1996. 'A Test for Payment Card Biases'. *Journal of Environmental Economics and Management* 31:178–85. https://doi.org/10.1006/jeem.1996.0039.

Ryan, Mandy, David A. Scott, and Cam Donaldson. 2004. 'Valuing Health Care Using Willingness to Pay: A Comparison of the Payment Card and Dichotomous Choice Methods'. *Journal of Health Economics* 23 (2): 237–58. https://doi.org/10.1016/j.jhealeco.2003.09.003.

Sandorf, Erlend Dancke, Margrethe Aanesen, and Ståle Navrud. 2016. 'Valuing Unfamiliar and Complex Environmental Goods: A Comparison of Valuation Workshops and Internet Panel Surveys with Videos'. *Ecological Economics* 129 (September):50–61. https://doi.org/10.1016/j.ecolecon.2016.06.008.

Schläpfer, Felix, and Michael Getzner. 2020. 'Beyond Current Guidelines: A Proposal for Bringing Behavioral Economics to the Design and Analysis of Stated Preference Surveys'. *Ecological Economics* 176 (October). https://doi.org/10.1016/j.ecolecon.2020.106720.

Schwarz, Norbert. 1997. 'Cognition, Communication, and Survey Measurement: Some Implications for Contingent Valuation Surveys'. In *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of*

Contingent Valuation Methods, edited by Raymond J. Kopp, Werner W. Pommerehne, and Norbert Schwarz, 165–88. Boston, MA: Kluwer.

Shackley, Phil, and Simon Dixon. 2014. 'The Random Card Sort Method and Respondent Certainty in Contingent Valuation: An Exploratory Investigation of Range Bias'. *Health Economics (United Kingdom)* 23 (10): 1213–23. https://doi.org/10.1002/hec.2980.

Skeie, M.A., H. Lindhjem, S. Skjeflo, and S. Navrud. 2019. 'Smartphone and Tablet Effects in Contingent Valuation Web Surveys – No Reason to Worry?' *Ecological Economics* 165. https://doi.org/10.1016/j.ecolecon.2019.106390.

Slovic, Paul. 1995. 'The Construction of Preference'. *American Psychologist* 50 (5): 364–71. https://doi.org/10.1037/0003-066X.50.5.364.

Smith, Richard D. 2006. 'It's Not Just What You Do, It's the Way That You Do It: The Effect of Different Payment Card Formats and Survey Administration on Willingness to Pay for Health Gain'. *Health Economics* 15 (3): 281–93. https://doi.org/10.1002/hec.1055.

Soeteman, Lotte, Job van Exel, and Ana Bobinac. 2017. 'The Impact of the Design of Payment Scales on the Willingness to Pay for Health Gains'. *European Journal of Health Economics* 18 (6): 743–60. https://doi.org/10.1007/s10198-016-0825-y.

Tversky, Amos, and Daniel Kahneman. 1974. 'Judgment under Uncertainty: Heuristics and Biases'. *Science* 185 (4157): 1124–31. https://doi.org/10.1126/science.185.4157.112.

Ungemach, Christoph, Adrian R. Camilleri, Eric J. Johnson, Richard P. Larrick, and Elke U. Weber. 2018. 'Translated Attributes as Choice Architecture: Aligning Objectives and Choices through Decision Signposts'. *Management Science* 64 (5): 2445–59. https://doi.org/10.1287/mnsc.2016.2703.

Veisten, Knut, Hans Fredrik Hoen, Ståle Navrud, and Jon Strand. 2004. 'Scope Insensitivity in Contingent Valuation of Complex Environmental Amenities'. *Journal of Environmental Management* 73:317–31. https://doi.org/10.1016/j.jenvman.2004.07.008.

Verbeek, Marno. 2017. *A Guide to Modern Econometrics*. 5th ed. Hoboken, NJ: John Wiley & Sons, Inc.

Veronesi, Marcella, Anna Alberini, and Joseph C. 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 (2): 199–215. https://doi.org/10.1007/s10640-010-9430-1.

Voltaire, Louinord, P. Wilner Jeanty, Claudio Pirrone, and Pierre Alexandre Mahieu. 2019. 'A Convergent Validity Test within the Payment Card Format Using Simulation Techniques'. *Applied Economics* 51 (34): 3770–86. https://doi.org/10.1080/00036846.2019.1584382.

Vossler, Christian A., and J. Scott Holladay. 2018. ‘Alternative Value Elicitation Formats in Contingent Valuation: Mechanism Design and Convergent Validity’. *Journal of Public Economics* 165 (September):133–45. https://doi.org/10.1016/j.jpubeco.2018.07.004.

Vossler, Christian A, and Ewa Zawojska. 2020. ‘Behavioral Drivers or Economic Incentives? Toward a Better Understanding of Elicitation Effects in Stated Preference Studies’. *Journal of the Association of Environmental and Resource Economists* 7 (2): 279–303. https://doi.org/10.1086/706645.

Wang, Hua. 1997. ‘Treatment of “‘Don’t-Know’” Responses in Contingent Valuation Surveys: A Random Valuation Model’. *Journal of Environmental Economics and Management* 32:219–32. https://doi.org/10.1006/jeem.1996.0965.

Wang, Hua, Jie He, Yoonhee Kim, and Takuya Kamata. 2013. ‘Willingness-to-Pay for Water Quality Improvements in Chinese Rivers: An Empirical Test on the Ordering Effects of Multiple-Bounded Discrete Choices’. *Journal of Environmental Management* 131 (December):256–69. https://doi.org/10.1016/j.jenvman.2013.07.034.

Wang, Hua, and Dale Whittington. 2005. ‘Measuring Individuals’ Valuation Distributions Using a Stochastic Payment Card Approach’. *Ecological Economics* 55 (2): 143–54. https://doi.org/10.1016/j.ecolecon.2004.11.011.

Welsh, Michael P., and Gregory L. Poe. 1998. ‘Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach’. *Journal of Environmental Economics and Management* 36 (2). https://doi.org/10.1006/jeem.1998.1043.

Whitehead, John C. 2002. ‘Incentive Incompatibility and Starting-Point Bias in Iterative Valuation Questions’. *Land Economics* 78 (2): 285–97. https://doi.org/10.2307/3147274.

Zawojska, Ewa, Malte Welling, and Julian Sagebiel. 2024. ‘Effects of Varying the Location of Perceived Consequentiality Elicitation in a Discrete Choice Experiment Survey’. *Land Economics* 100 (2): 223–44. https://doi.org/10.3368/le.100.2.040320-0048R1.

Zhang, Jing, and Wiktor L Adamowicz. 2011. ‘Unraveling the Choice Format Effect: A Context-Dependent Random Utility Model’. *Land Economics* 87 (4): 730–43. https://doi.org/10.3368/le.87.4.730.

Tables

Table 1. Data cleaning based on commonly applied exclusion criteria in contingent valuation studies for protest responses, inconsistent responses, and WTP outliers.

	Control	Treatment 1	Treatment 2
Observations in the gross sample	1,177	1,272	1,258
Protest responses[a]	144	190	199
Inconsistent responses[b]	142	110	119
WTP outliers[c]	5	7	9
Observations in the net sample for analysis	886	965	931

Notes:
a: *Zero WTP response for self-reported reason classified as a protest in a follow-up question.*
b: *WTP responses that are not weakly monotonically increasing in ecosystem service damage, e.g., if stated WTP for a given ecosystem service damage level is lower than that of a preceding, lower level of ecosystem service damage. In addition, 11, 28, and 13 respondents classified as both protest and inconsistent are counted under protest only for the Control, Treatment 1, and Treatment 2 conditions, respectively.*
c: *Stated WTP exceeding 5% of personal gross annual income, for any ecosystem service damage level. 2, 2, and 3 respondents classified as both inconsistent and outlier are counted under inconsistent only for the Control, Treatment 1, and Treatment 2 conditions, respectively.*

Table 2. Overview of independent variables.

Variable	Description	*N*	Mean	SD	Min.	Max.	Category	Frequency	Proportion
PAYMENTSCALE	Categorical, experimental condition	2782					Control	886	0.32
							Treatment 1	965	0.35
							Treatment 2	931	0.33
INCOMELEVEL21	Categorical, gross personal income in 2021	2362					<NOK 400,000	567	0.24
							[NOK 400,000, NOK 800,000)	1381	0.58
							[NOK 800,000, NOK 1,200,000)	346	0.15
							>NOK 1,200,000	68	0.03
AGE	Age of respondent	2782	56.3	16.5	17	91			
AGESQ	Age of respondent squared	2782	3440.9	1792.3	289	8281			
HIGHED	Dummy, 1 if completed graduate level education (MA/PhD)	2782	0.28	0.45	0	1			
MALE	Dummy, 1 if male respondent	2782	0.50	0.50	0	1			
MEMBER	Dummy, 1 if member of recreational and/or environmental organizations	2782	0.29	0.45	0	1			
PROXIMITY	Categorical, three categories of household proximity to the sea	2597					>100 km	230	0.09
							[10 km, 100 km]	745	0.29
							<10 km	1622	0.62
PREVIOUSEXP	Dummy, 1 if previous experience with the environmental insult of the contingent valuation scenarios	2608	0.23	0.42	0	1			
SMARTPHONE	Dummy, 1 if using a smartphone as response platform	2782	0.27	0.44	0	1			

Note: N denotes the number of observations in the net sample for analysis after implementation of standard exclusion criteria in contingent valuation studies on the gross sample, as described in Section 3.

Table 3. Descriptive statistics for WTP (selected payment scale bid) in NOK to avoid “small” (S), “medium” (M), “large” (L), and “very large” (XL) ecosystem service damage. Mean values excluding open-ended WTP responses (when WTP exceeds the highest payment scale bid) are given in parenthesis.

	WTP_S			WTP_M			WTP_L			WTP_{XL}		
	C	T1	T2	C	T1	T2	C	T1	T2	C	T1	T2
Mean	661 (661)	777 (777)	944 (944)	998 (980)	1336 (1219)	1395 (1395)	1497 (1418)	2004 (1852)	2060 (1999)	2208 (1948)	2774 (2392)	2703 (2427)
Median	400	500	600	500	600	1,000	900	1,000	1,200	1,200	1,000	1,400
Min.	0	0	0	0	0	0	0	0	0	0	0	0
Max.	12,000	10,000	10,000	15,000	100,000	10,000	20,000	100,000	20,000	40,000	100,000	30,000

Table 4. Interval regression analysis of effects of payment scale conditions on WTP in NOK to avoid “small” (S), “medium” (M), “large” (L), and “very large” (XL) ecosystem service damage.

	WTP_S	WTP_M	WTP_L	WTP_{XL}
PAYMENTSCALE				
Control (baseline)				
Treatment 1	189.785***	344.906***	561.75***	668.533***
	(71.361)	(94.782)	(129.863)	(163.567)
Treatment 2	326.783***	432.604***	627.305***	565.96***
	(71.696)	(95.103)	(130.097)	(164.477)
INCOMELEVEL21				
[0, NOK400,000) (baseline)				
[NOK400,000, NOK800,000)	218.282***	335.707***	494.507***	756.381***
	(73.735)	(97.975)	(134.216)	(169.181)
[NOK800,000, NOK1,200,000)	318.747***	428.587***	659.293***	954.966***
	(101.627)	(134.488)	(184.148)	(232.45)
NOK1,200,000 or above	57.809	180.768	699.165**	1609.822***
	(181.312)	(243.206)	(332.462)	(422.574)
AGE	10.423	24.371	48.451**	43.857
	(12.283)	(16.295)	(22.278)	(28.002)
AGESQ	-.04	-.145	-.331	-.288
	(.113)	(.15)	(.205)	(.257)
HIGHED	66.389	117.005	267.729**	229.692
	(66.676)	(88.47)	(121.053)	(153.118)
MALE	-92.676	-131.062	-78.14	-27.559
	(60.444)	(80.212)	(109.676)	(138.373)
MEMBER	128.501**	299.985***	439.562***	760.217***
	(63.281)	(83.998)	(114.966)	(145.418)
PROXIMITY				
> 100 km. (baseline)				
[10 km., 100 km.]	25.711	112.977	64.9	-85.911
	(113.611)	(152.047)	(207.794)	(260.97)
< 10 km.	98.87	226.551	343.509*	175.829
	(107.218)	(143.556)	(196.13)	(246.279)
PREVIOUSEXP	228.533***	357.21***	570.503***	658.525***
	(67.351)	(89.519)	(122.716)	(156.136)
SMARTPHONE	-19.231	64.065	94.951	184.774
	(68.788)	(91.538)	(125.41)	(158.911)
Constant	-8.76	-346.356	-916.2	-374.545
	(334.643)	(445.143)	(608.473)	(764.728)
Observations	1932	1911	1903	1864

*Note: Standard errors are in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01.*

Table 5. Regression analysis of effects of payment scale conditions on WTP in NOK to avoid "small" (S), "medium" (M), "large" (L) and "very large" (XL) ecosystem service damage. Tobit estimation, with the selected payment scale bid as point estimates of WTP. Marginal effects for censored WTP.

	WTP_S	WTP_M	WTP_L	WTP_XL
PAYMENTSCALE				
Control (baseline)				
Treatment 1	92.18*	187.20***	328.97***	420.41***
	(49.77)	(69.17)	(99.58)	(144.51)
Treatment 2	197.88***	272.04***	402.93***	340.21**
	(51.37)	(70.51)	(100.68)	(144.83)
INCOMELEVEL21				
[0, NOK400,000) (baseline)				
[NOK400,000, NOK800,000)	165.91***	249.14***	380.83***	618.42***
	(50.60)	(70.11)	(100.44)	(142.64)
[NOK800,000, NOK1,200,000)	242.50***	329.07***	507.56***	886.05***
	(73.13)	(100.15)	(143.59)	(207.03)
NOK1,200,000 or above	63.45	134.32	479.66*	1191.06***
	(124.95)	(176.71)	(265.32)	(401.05)
AGE	5.42	15.56	36.01**	31.29
	(8.81)	(12.15)	(17.41)	(24.96)
AGESQ	-0.01	-0.09	-0.26	-0.23
	(0.08)	(0.11)	(0.16)	(0.23)
HIGHED	50.37	83.72	214.48**	227.44*
	(47.72)	(65.90)	(94.42)	(136.24)
MALE	-80.84*	-105.60*	-48.66	-15.91
	(43.34)	(59.87)	(85.82)	(123.45)
MEMBER	97.77**	226.23***	343.25***	525.87***
	(45.26)	(62.48)	(89.63)	(129.19)
PROXIMITY				
> 100 km. (baseline)				
[10 km., 100 km.]	71.06	148.99	127.58	8.20
	(78.95)	(108.77)	(155.96)	(229.04)
< 10 km.	124.31*	236.23**	354.00**	244.96
	(74.51)	(102.56)	(147.71)	(217.09)
PREVIOUSEXP	169.73***	277.14***	436.64***	611.69***
	(48.10)	(66.52)	(95.53)	(138.54)
SMARTPHONE	-12.24	39.68	68.31	97.18
	(49.27)	(68.28)	(98.04)	(141.64)
Observations	1932	1911	1903	1864

*Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

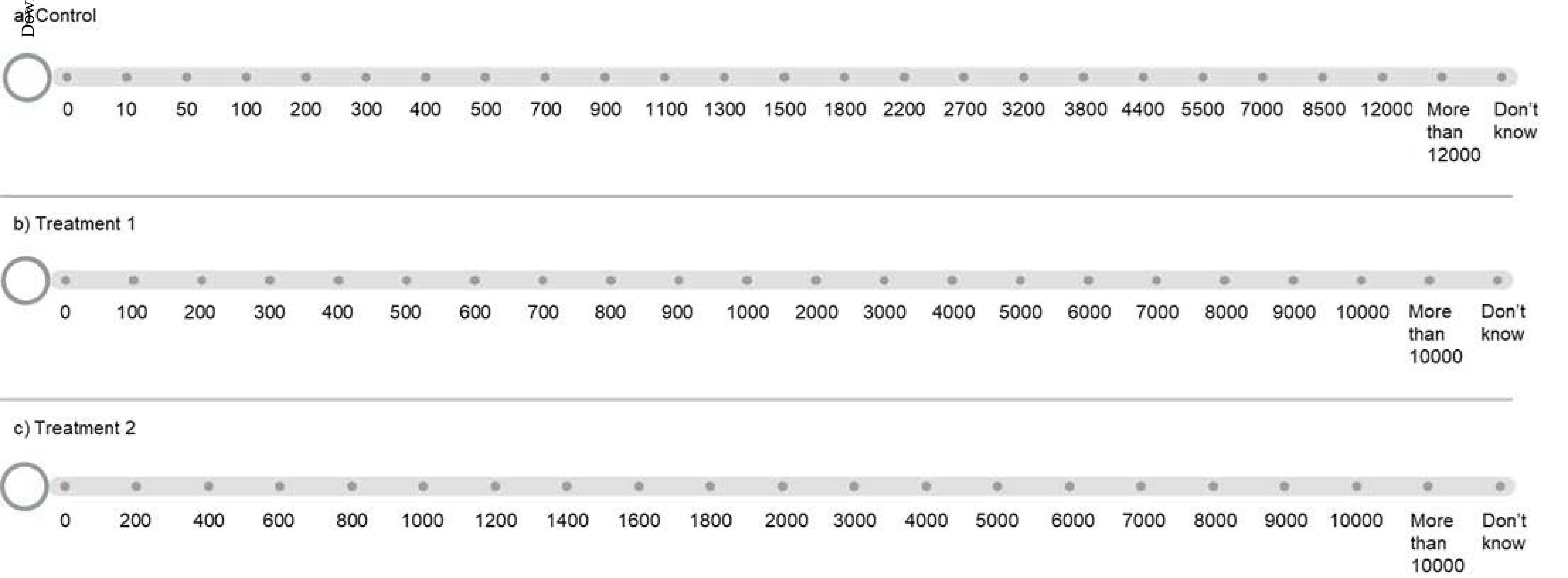
Figures

Figure 1. Experimental payment card conditions in the contingent valuation survey: a) Control, b) Treatment 1, and c) Treatment 2.

Figure 2. Ecosystem service damage associated with four oil spill levels in the contingent valuation survey.

---

[1] Single binary choice, double-bounded binary choice, open-ended response, and payment card formats.
[2] The slider bar design is frequently used in web surveys to elicit responses in various domains (e.g., Maineri, Bison, and Luijkx 2021).
[3] Typically, at least 30% of respondents conduct the survey on a smartphone or tablet (Skeie et al., 2019).
[4] In the two Treatment conditions, less than 5% of respondents stating their WTP in the payment scale used the open-ended response opportunity. For the rest, maximum WTP either corresponded to the chosen bid or they did not take their time to consider or enter an exact amount within the implied interval. Therefore, this feature is not given further attention.
[5] There is no standard threshold for WTP outliers in contingent valuation. Other thresholds frequently used include WTP > 10% of income (e.g., Mitchell and Carson 1989), WTP > 2% of income (e.g., Carson, Wilks, and Imber 1994; Veisten et al. 2004), and WTP > 3 × Interquartile range (IQR) (e.g., Frey and Pirscher 2018).
[6] Ecological validity refers to the degree to which findings from a particular experimental context is transferrable to the typical “real-world” situation in which the outcome manifests (Holleman et al. 2020).
[7] Sampling validity denotes the extent to which the sampling process of a study yields a representative sample of the population to which results are generalized (McEwan 2020).
[8] The affected population includes all households in the Norwegian counties of Akershus (326 435), Buskerud (125 196), Vestfold (119 458), and Østfold (145 121). Counts as of 2022.
[9] In addition to all households included for the “small” ES damage level, the affected population includes all households in the counties of Oslo (333 251), Telemark (82 843), Agder (148 159), and Innlandet (184 431). Counts as of 2022.
[10] Only 1.6% and 4.6% of the respondents selected the endpoint bids of NOK12,000 and NOK10,000, respectively. Further, only 1.7% of respondents provided an open-ended answer after selecting the option “more than [NOK12,000 or NOK10,000]” across all payment scale conditions and elicitations.
[11] See, e.g., Ungemach et al. (2018).


a) Control
0
10
50
100
200
300
400
500
700
900
1100
1300
1500
1800
2200
2700
3200
3800
4400
5500
7000
8500
12000
More than 12000
Don't know
b) Treatment 1
0
100
200
300
400
500
600
700
800
900
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
More than 10000
Don't know
c) Treatment 2
0
200
400
600
800
1000
1200
1400
1600
1800
2000
3000
4000
5000
6000
7000
8000
9000
10000
More than 10000
Don't know