Multiple Price Lists for Willingness to Pay Elicitation

B. Kelsey Jack
Kathryn McDermott
Anja Sautmann

Abstract

Multiple price lists are a convenient tool to elicit willingness to pay in surveys and experiments, but choice patterns such as "multiple switching" and "never switching" indicate high error rates. Existing measurement approaches often do not provide accurate standard errors and cannot correct for bias due to framing and order effects. This paper proposes to combine a randomization approach with a random-effects latent utility model to detect bias and account for error. Data from a choice experiment in South Africa shows that significant order effects exist which, if uncorrected, would lead to distorted conclusions about subjects' preferences. Templates are provided to create a multiple price list survey instrument in SurveyCTO and analyze the resulting data using the proposed methods.

Multiple Price Lists for Willingness to Pay Elicitation[*]

B. Kelsey Jack   Kathryn McDermott   Anja Sautmann

---

1 Introduction

Multiple price lists (MPL) are a widely used and convenient tool to elicit subject prefer-ences as part of a survey or experiment. MPL elicitation combines ease of implementation with detailed information about preferences, which makes the method attractive for field settings. Researchers have used it to elicit willingness to pay (WTP) for a diverse set of goods and services, from timed rewards to cookstoves.[1] Willingness to pay (WTP) here refers to the subject's (relative) preferences between two options expressed in a common unit, such as money. In an MPL, preferences may be elicited by asking repeated questions that allow the subject to "buy" or "sell" indivisible goods at various prices, or to choose between divisible goods at different quantities, thus establishing the subject's (relative) valuation of these goods. The use of MPL in WTP elicitation is not new; for example, Kahneman et al. (1990) employ it in their seminal laboratory study on the endowment effect, involving subjects who buy and sell everyday items such as pens or mugs. Since then, the literature has grown to include both methodological studies (e.g. Andersen et al., 2007) and applied measurement (e.g. Bursztyn et al., 2018).[2]

Researchers in applied microeconomics and development economics typically measure WTP in the context of a larger project and therefore look to the existing literature for the best measurement method. Perhaps for this reason, there has been renewed interest in validating WTP elicitation methods. Several recent papers have focused on the broader question of comparing methods, such as take-it-or-leave-it (TIOLI) vs. the Becker-deGroot-Marschak (BDM) random price mechanism (Berry et al., 2020; Cole et al., 2020). MPLs are situated somewhere between these methodologies: they provide an incentive-compatible elicitation method that delivers more information than TIOLI, but is easier to understand and more widely applicable than BDM.[3]

Among studies that use MPL to elicit WTP, data collection instruments and estima-tion methods vary considerably. Much of this variation is driven by a set of common

---

[1]See Appendix B for an illustrative list of papers and Burchardi et al. (2021) for an overview.

[2]MPL are also commonly used for the elicitation of other preference parameters, such as time prefer-ences and risk aversion (e.g. Andersen et al. (2008)), and much of the methodological research on MPL has been done in that literature. A key difference is that risk aversion and time preference elicitation are primarily used to establish ordinal preferences (through the parameters in a utility function), whereas in WTP elicitation, the cardinal valuation is often a key research interest.

[3]Many authors treat MPL between a single item and money as a special case of BDM (e.g., Cole and Fernando (2020); Maffioli et al. (2020)). We think of MPL more broadly as a format for eliciting preferences between any two different options, which includes that one or both of these options are money. The key feature of BDM is that the final price is selected randomly, which makes the mechanism incentive compatible with stating the true WTP, and MPL share this feature. Thus, MPL provide a modified format to implement BDM.

challenges that the existing approaches have only partially addressed. First, many subjects exhibit "multiple switching behavior" (MSB), meaning they make inconsistent choices that are difficult to interpret, by switching between "buying" vs. "not buying" multiple times as prices change monotonically (Yu et al., 2021). Based on these decisions, their WTP may lie in a very wide interval or not be recoverable at all.[4] Second, many subjects exhibit "never-switching behavior" (NSB), meaning they choose the same option throughout the MPL, which may imply a very wide interval for the WTP including potentially $+\infty$ or $-\infty$. Sometimes, one of these choices is inconsistent because the selected option is strictly dominated.[5] Last but not least, even consistent responses in MPLs only provide interval identification of WTP.

We begin by introducing a framework aimed at understanding choice error in WTP elicitation, and its potential to add noise to WTP estimates. We argue that MPL can be an effective tool for measuring WTP because of the repeated choice data from the same subject that permits learning about error. However, due to their format of repeated parallel questions, MPL may also be vulnerable to framing and order effects. These create bias because they introduce a correlation between the error and the (change in) value difference between the two options, which is used in an MPL to identify WTP. We argue that many widely used MPL designs do not accurately measure error and are unable to prevent or even detect order effects. Worse, some designs may introduce additional bias. Moreover, most estimation approaches do not model choice error, with implications for the estimated variation and standard errors in WTP.

We therefore propose two measures to improve how MPL-based WTP elicitation incorporates choice error and bias. The first measure is to randomize how subjects view choices in a way designed to address potential framing effects. We propose to vary both the order of binary choices within MPL, so that the value difference between options – aka the "price" – is either ascending or descending,[6] and the order of options within each binary choice, that is, whether an option appears first or second/on the left or right.[7] This

---

[4]Take the case of a "reverse switch" where the subject declines to buy at all low prices offered, but then switches and agrees at a series of higher prices. Formally, they express that their WTP is below x but also above $x + t$, with $t \gg 0$.

[5]Dominated choices are often included in the MPL as a test of rationality.

[6]As we discuss in more detail in section 3.1, we do not propose full randomization within MPL to reduce subjects' cognitive load; see also Andersen et al. (2007) for the same argument.

[7]We think here in particular of visual presentations of the MPL, e.g. on paper or on a screen, but the order in which the two options are listed may also matter in orally administered surveys. When subjects make yes/no decisions about a good at different prices, the researcher may vary order within choice by varying whether the subject is asked to buy or sell a good, or whether the "yes" or "no" answer option is presented first.

minimizes net bias and allows us to distinguish anchoring effects that arise from simply "repeating" the same decision from explicit choices that reveal true WTP. The second measure is to model each binary choice in the MPL with a latent utility model and apply an appropriately scaled random-effects probit estimation to estimate WTP. This approach accommodates inconsistency (MSB) as well as dominated choices, and thus avoids selection biases that arise when dropping or recoding MSB or NSB choices. In addition, it provides identification of average WTP, dispersion of individual WTP, and errors in subjects' choices, and allows us to explicitly estimate biases that arise from order effects.[8] This ability to learn about both the distribution and mean of individual errors is a key advantage of the MPL method relative to alternative preference elicitation methods.

We show that both non-systematic errors and systematic biases play a significant role in individual choices, using a WTP elicitation example from South Africa. This finding is in line with the literature; in a survey of MPL studies, 17.1 percent of observations included inconsistencies (Filippin and Crosetto, 2016; Crosetto and Filippin, 2016). Research in populations with low literacy and numeracy may be particularly affected: for example, Dave et al. (2010) show that subjects with low math ability make significantly more inconsistent choices. Moreover, while most researchers do not randomize the order in which MPL questions are asked, others have found significant order effects (e.g. Channa et al., 2021). In our application, the order of choices within MPL is an important source of bias in measured WTP, but the order of options within choice has little effect on reported valuations.

We compare our WTP estimates with those we would have obtained using MPL design and estimation approaches found in the literature, and show important differences. First, WTP estimates obtained from non-randomized MPL cannot differentiate true WTP from existing order bias, and some MPL even create new sources of bias (such as MPL elicitation that stops after the first "switch"). Second, not accounting for choice error alters the estimated outcome variance. Together, these issues lead to markedly different conclusions about the mean and dispersion of WTP in the study population.

To facilitate use of our proposed MPL design and estimation approach, we provide a template WTP elicitation module implemented in the survey software SurveyCTO that carries out order randomization within the MPL. In addition, we provide template analysis files in Stata that carry out the random-effects latent utility estimation. The template

---

[8]Note that this methodology can be applied to measures of the sample average WTP, or to subgroup-level measures of WTP, but not WTP at the individual level. Based on the high rate of choice errors in typical MPL data from LMIC contexts, we caution against approaches that calculate an individual-specific WTP from MPL data, for use, for example, as an explanatory variable.

accommodates order randomization as well as a set of other features identified as good practice for MPL design from the authors' own research and the literature. These features include practice MPL rounds, constant maximal values in each binary choice within MPL, focusing WTP intervals on points of interest for hypotheses tests, and others. Many of these can also be implemented in pen and paper or verbally elicited MPL.

Our paper contributes to the methodological literature on WTP elicitation in field applications (e.g., Berry et al. (2020); Burchardi et al. (2021)). We focus on multiple price lists, an increasingly common elicitation approach in development economics. MPL are popular because they hold the promise of both straightforward implementation for researchers and ease of comprehension for subjects. However, we show that choice error is an important feature of MPL data, and likely any WTP data. When combined with specific features of data collection or analysis, this error has the potential to introduce bias in measured WTP. While others have discussed framing effects in MPL data (e.g., Andersen et al. (2007)), order effects have been largely overlooked to date. We draw three main conclusions: first, in populations where choice error is frequent, we caution against the use of some of the most common methods used in the design and analysis of MPL. Second, researchers using MPL or interpreting MPL data should pay special attention to the potential for order bias. Third, with the right MPL design and estimation approach, we believe that MPL continue to be a valuable tool. Our proposed two-pronged method of order randomization and latent utility estimation provides researchers and practitioners with a unified approach to diagnosing and mitigating bias while correctly accounting for choice error. Our Stata and SurveyCTO templates facilitate the measurement of WTP using MPL.

The next section describes WTP elicitation via MPL. It discusses common patterns in MPL data, presents a framework to think about error in such data, and summarizes how the existing literature has addressed MSB, NSB and interval identification. Section 3 proposes to combine order randomization with a random utility model for MPL choices and uses data from an experiment in Cape Town, South Africa to show how this approach affects measured WTP. We demonstrate in particular that order biases can significantly affect elicited WTP. We also describe how researchers might apply our framework to investigate other sources of bias that arise through MPL design decisions. In section 4, we discuss features of MPLs for WTP elicitation that we identify as useful for improving data quality. These features are implemented in an accompanying template for implementing an MPL in SurveyCTO and estimating WTP from the data, described in detail in a technical appendix (Appendix S2).

4

2 Multiple price lists for willingness to pay elicitation

Many experimental preference elicitation procedures use some form of list experiment. We focus here on MPL (Multiple Price Lists) that elicit the subject's relative preference for an option A over a second option B expressed in a common (monetary) value, that is, the subject's willingness to pay (WTP) to receive option A over option B. However, MPLs can be and frequently are used for other purposes, such as measuring risk or time preferences, and much of our analysis and discussion applies.

In a WTP MPL, the subject makes repeated binary choices between two options, where the monetary value of these options is varied systematically. To fix ideas, we reproduced four examples of MPL experiments from the literature that fit this description in Figure 1. These examples show a variety of designs, and illustrate some key properties of MPL. Our analysis below draws on many other papers as well, and Appendix B provides an overview of 23 papers that use MPL to measure WTP, with a focus on incentivized MPL from low and middle income countries (LMIC). The wide range of contexts and topics covered demonstrate the versatility of the method; the large number of papers since 2019 show that MPL is increasingly used in applied research to measure WTP. We caution that this list is not exhaustive and other papers have influenced our work, including the large literature on risk aversion and time preference elicitation. However, the papers in Appendix Table B showcase a diverse set of high-quality papers directly related to WTP measurement, providing concrete examples to researchers who are designing an MPL experiment.

We use the term "option" to refer to the components that remain the same within the MPL. An option can be a specific good, a task, or a lottery; but also the format or time of delivery of the same commodity, including money, for example when eliciting time preferences. In each row of the MPL, the respondent makes a choice between two options, with varying monetary values associated with each option. The inference on the subject's willingness to pay comes from the value difference between the two options in each row of the MPL. For example, in the MPL from Allcott and Kessler (2019) in the top left of Figure 1, one option is to receive four home energy reports (for residential electricity users), and the other option is not to receive reports. The associated cash prize varies between $1 and $10 for each option; from high to low when choosing the report (left) but low to high when choosing no report (right). As a result, the difference in monetary value between the two options varies between -9 and 9 US dollars.

Figure 1: Examples of MPL-style WTP elicitation in published papers.

Which would you prefer?				
1	$10	4 more Home Energy Reports PLUS a $10 check	$1	A $1 check
2	$10	4 more Home Energy Reports PLUS a $10 check	$5	A $5 check
3	$10	4 more Home Energy Reports PLUS a $10 check	$9	A $9 check
4	$10	4 more Home Energy Reports PLUS a $10 check	$10	A $10 check
5	$9	4 more Home Energy Reports PLUS a $9 check	$10	A $10 check
6	$5	4 more Home Energy Reports PLUS a $5 check	$10	A $10 check
7	$1	4 more Home Energy Reports PLUS a $1 check	$10	A $10 check

Start with the 500 Sh ticket, and show it to the participant.

If the prize was either 500 Sh ANNOUNCED or 500 Sh NOT ANNOUNCED, which would you choose?

If the prize was either 500 Sh ANNOUNCED or 450 Sh NOT ANNOUNCED, which would you choose?

If the prize was either 500 Sh ANNOUNCED or 400 Sh NOT ANNOUNCED, which would you choose?

...

If the prize was either 500 Sh ANNOUNCED or 150 Sh NOT ANNOUNCED, which would you choose?

Tick if you are willing to buy at the price in each row and cross "x" if you are not willing to buy.

		Organic	Conventional
1	50		
2	100		
3	150		
...	...		
17	850		
18	900		
19	1000		

SurveyCTO questionnaire			
1	Would you buy the card for 0 Shillings (for free)?	yes	no
2	Would you buy the card for 200 Shillings?	yes	no
...	...	..	..
11	Would you buy the card for 2000 Shillings?	yes	no

Notes: Clockwise from top left: Allcott and Kessler (2019) sent out paper questionnaires to home energy users to measure preferences for receiving usage reports. In Squires (2021), a surveyor read out choices between receiving payments of different value either in public or in private, and varied the order of choices between subjects. In Burchardi et al. (2021), a surveyor read out yes/no questions about purchasing a voucher directly redeemable for money at increasing prices. Alphonce and Alfnes (2017) measured preferences for organic and conventional produce for tomato buyers at a market in Tanzania (prices in Tanzanian Shillings).

MPL can be used to carry out a variant of the Becker-deGroot-Marschak (BDM) mechanism, and many authors use the term BDM to refer to an MPL for WTP elicitation. In a "pure" BDM, subjects are asked to state their true value for an object, a random (purchase) price is drawn, and the purchase is realized if the stated value is higher than the price (the reverse for a "sale"). In the equivalent MPL, subjects are explicitly asked whether they are willing to buy (or sell) an object at a series of different prices, and a single row is drawn for implementation. For example, in the price experiment from Burchardi et al. (2021) in the bottom right of Figure 1, subjects are asked whether they would like to buy a voucher ("card") at prices starting at 0 and going up to 2000 Shillings.

In most MPL, the binary choices subjects make are ordered by either ascending or descending value difference. The order of binary choices, and, where applicable, the order of the options presented in each binary choice (e.g., on the right or left side of the list), do not vary within or between subjects.[9] On paper or on a screen, subjects may see all choices in front of them at the same time. When administered via a computerized questionnaire or oral instruction by the enumerator, choices for each row may be elicited before revealing the next row. At the end of the experiment, one binary choice is usually chosen for implementation, and the subject's decision in that binary choice is realized.

In many research projects, MPLs are one element of a larger data collection effort. Nonetheless, since the value subjects attach to different options is often a key factor of interest, researchers put a lot of thought into the design of these MPLs. For example, the original MPL in Allcott and Kessler (2019), reproduced in Figure 1, was a paper questionnaire that used a variety of visual aids to help subjects make their decisions. Many authors report extensive piloting or test different MPL formats, and survey questionnaires often check subject comprehension, allow revisions, or provide "practice" MPLs and other tools to improve measurement. The detailed supplemental material in Berkouwer and Dean (2021) offers a nice example, and we encourage the reader to explore the papers listed in Appendix B.

---

[9]The MPL implemented in Squires (2021), shown on the top right in Figure 1, is one exception. In his study of WTP for private (vs. public) income, he randomized the order of binary choices: some subjects saw the choice including 500 Sh first, and some saw the choice including 150 Sh first.

2.1 Multiple switching and never-switching: Common issues with MPL data

Despite their widespread use, there are a number of issues with MPLs that many, if not most, researchers have encountered. One common issue is inconsistent choices, often termed multiple switching behavior (MSB). Take a (hypothetical) subject who chooses "no report" in row 2, but "report" in row 3 of the Allcott and Kessler (2019) MPL. This subject has the preferences ($5, no report) $\succcurlyeq$ ($10, report) $\succcurlyeq$ ($9, no report) and, by transitivity, could be inferred to prefer $5 to $9. The term "multiple switching" refers to the fact that consistent choices require that the subject switches between the two options at most once, from the MPL option with decreasing (relative) value to the option with increasing value.

In addition to MSB, many subjects exhibit never-switching behavior (NSB). NSB occurs when a subject chooses the same option through the entire price list. Since the researchers typically vary the value difference between the two options across the full support of a reasonable distribution, this behavior often appears implausible. For example, take a subject in Alphonce and Alfnes (2017) (see Figure 1) who chooses to buy conventional tomatoes at all prices up to TZS 1000 – this would be implausible given that the market price at the time of the experiment ranged between TZS 300 and 400. In the MPL from Burchardi et al. (2021) (also shown in Figure 1), NSB would be an even clearer indicator of irrational choices, since valuations are elicited for a "card" that is directly redeemable for its face value of UGX 1400 in cash, and all subjects should be willing to buy the card for a price below 1400 but not for a price above; in fact, the authors use this as a rationality test when comparing different elicitation methods. In general, however, a problem for the measurement of subjective preferences is that the researcher cannot distinguish between NSB that is the expression of a very strong preference for one of the options, and NSB that is the result of an error.

A survey of 54 published risk elicitation studies employing MPL designs found that 17.1 percent of observations included inconsistencies in the form of MSB or subjects picking dominated options (Filippin and Crosetto, 2016; Crosetto and Filippin, 2016).[10] In WTP elicitation where the trade is realized later, subjects often ex post decline the agreed-upon price: Maffioli et al. (2020) report a high rate of reneging on MPL outcomes,

---

[10]A similar systematic review of the MPL for WTP literature is not available. Of the papers we surveyed, a subset of which are shown in Appendix B, MSB was rarely reported and was often suppressed by instrument design, as we discuss in greater detail in Section 2.3. NSB was more commonly reported, and ranged from 1-2 percent to 74 percent of subjects.

and summarize a number of other studies that report reneging rates above 10 percent. This suggests that individual decisions in the MPL may not always reflect the subject's relative valuation of the two options, and these errors cannot always be detected and corrected. In the best case, this adds noise around the true willingness to pay. In the worst case, it introduces systematic bias.

2.2   Measurement error in MPL: A framework

We start with a simple framework to think about confounders and measurement error in a WTP MPL. We assume that utility is additive in monetary value. Thus, option A provides subject i with utility $U_{iA} + v_{At}$ and option B provides $U_{iB} + v_{Bt}$, where t denotes the binary choice the subject is facing, U_{ij} is the utility from option j, and v_{jt} is the monetary value associated with j in t. We assume that the researcher is interested in the mean (or other moments) of the distribution of true underlying differences in utility between options, expressed in money. We call an individual difference the subject's willingness to pay for option A: $WTP_i = U_{iA} - U_{iB}$.

A typical MPL offers T binary choices between the same two options with different associated values v_{At} and v_{Bt}, $t = 1, \ldots, T$. If subjects never make mistakes and there are no outside factors that influence choices, these binary decisions provide interval identification for the WTP, because subject i will choose option A if

$$U_{iA} + v_{At} > U_{iB} + v_{Bt}$$
$$\text{or} \quad WTP_i > v_{Bt} - v_{At}, \tag{2.1}$$

option B if
$$WTP_i < v_{Bt} - v_{At},$$

and either A or B if she is indifferent. This means the WTP must lie in the closed interval between the largest value of $v_{Bt} - v_{At}$ where option A is chosen and the smallest $v_{Bt} - v_{At}$ where option B is chosen.

A first observation is that even without any errors, WTP is never point identified, and a distribution over the identified interval must be assumed. Moreover, in the NSB case, the WTP is either in the highest or lowest possible range, but the interval is unbounded: if option A is chosen for all t, we have that $WTP \in [\max_t(v_{Bt} - v_{At}), \infty)$ and if option B is always chosen, we have $WTP \in (-\infty, \min_t(v_{Bt} - v_{At})]$. Decisions about where in these intervals the WTP of never-switchers is likely to lie can have significant effects on

the distribution of elicited WTP.

Of course, some of the patterns discussed earlier indicate that there likely *is* error in individual choices; such as MSB, dominated choices (e.g., choosing not to buy a voucher that can be exchanged for cash), and implausible WTP distributions, such as a mass of never-switchers at both extremes of the MPL. Formally, the subject will weakly prefer option A if

$$U_{iA} + v_{At} + \epsilon_{iAmt} \geq U_{iB} + v_{Bt} + \epsilon_{iBmt},$$

$$\text{or} \quad WTP_i \geq v_{Bt} - v_{At} + \epsilon_{iBmt} - \epsilon_{iAmt}$$

$$\equiv -\Delta_t + \epsilon_{iBmt} - \epsilon_{iAmt}.$$

The terms ϵ_{ijmt} describe any disturbance to the preference for option j in choice t of MPL m. Because of this term, the subject may sometimes choose option A even though $WTP_i < v_{Bt} - v_{At}$ and option B even though $WTP_i > v_{Bt} - v_{At}$. This could lead to multiple switching as well as randomly switching "too early" or "too late". Going forward, we will use the value difference term $\Delta_t = v_{At} - v_{Bt}$. This is the inverse of the difference in values on the right hand side above, and so subjects will tend to choose option A when Δ_t is large and option B when it is small (incl. negative).

Potential reasons for individual error, ϵ_{ijmt}, include inattention and incomprehension. For example, Yu et al. (2021) have shown in the case of risk preferences that greater attention reduces MSB in final MPL choices. Cason and Plott (2014) also provide evidence that unexpected behavior in a BDM occurs when subjects do not understand the payoff structure of the game created by the experiment, leading to behavior that looks irrational. Incomprehension may be a particular concern in populations with low literacy and numeracy. For example, Dave et al. (2010) have shown that subjects with a math ability score more than one standard deviation below the sample mean make inconsistent choices about four times more often than do other participants. On the other hand, Allcott and Kessler (2019) report only 2.4% inconsistent choices in a mail-in survey with US energy customers.

Note that Andersen et al. (2008) have argued that choices that look inconsistent simply indicate that the subject is indifferent between a range of options. Complete indifference should be a knife-edge case if the MPL is well constructed and choices are realized, but the subject may still care little about the offered payoffs compared to, for example, the time it takes to go through them carefully, and therefore choose not to incur the attentional

cost.[11] Whatever the exact causes for error may be – inattention, a lack of comprehension, or near-indifference – in the extreme, these confounders could lead to choices that are only weakly related to true WTP, if subjects treat the options in each binary choice essentially as "the same". MSB with a large number of switches may in fact be a sign that some subjects randomize deliberately in choices that feel repetitive.[12]

The presence of error has consequences for the interpretation of choice data in elicitation experiments. On the one hand, even when detectable errors occur, the subject's choices still hold information, because they are correlated with true preferences, but this information is more difficult to extract. On the other hand, a large number of choice errors may remain undetected when the subject's choices are internally consistent, and any given MPL may not identify a subject's WTP. Most alternative elicitation methods, such as TIOLI or BDM, do even worse in this respect: they cannot identify inconsistencies or likelihood of error at all because the subject makes only one choice. We think of the ability to learn about the distribution of the ϵ_{jimt} from MSB as one key advantage of the MPL method.[13]

MPL design and order biases in WTP. If the difference $\epsilon_{iBmt} - \epsilon_{iAmt}$ is symmetrically distributed around zero and uncorrelated with the systematic component of the subject's preferences, the errors will introduce noise into measured WTP. Additional issues arise if the error term is correlated with the value difference Δ_t or has a nonzero mean. In this case, the data may lead to a systematic over- or underestimation of WTP. As a stylized example, suppose that $\epsilon_{iBmt} - \epsilon_{iAmt}$ is small (negative) whenever $v_{Bt} - v_{At}$ is large, i.e. Δ_t is small. Then the subject may frequently choose option A even though the utility from B is higher, leading to an overestimate of WTP.

Biases often occur due to framing or anchoring effects, because they introduce systematic (rather than random) deviations from true preferences: framing means that subjects choose one option more often only due to the way the choices are presented, either because the framing affects choice errors, or because it directly influences preferences (e.g.,

---

[11]Indifference, or near indifference, could occur in hypothetical MPL, and in situations where the utility difference between options is very small, such as for subjects who strongly discount future payoffs and choose between two payments far in the future, or very risk averse subjects who make choices between two lotteries.

[12]Agranov and Ortoleva (2017) have shown that subjects exhibit stochastic choice in repeated identical decisions that directly follow each other.

[13]Others have previously pointed out both that ignoring measurement error in experimental preference elicitation can lead to biased estimates (e.g. Gillen et al. (2019)), and that detected measurement error may contain valuable information (Andersen et al., 2008).

Tversky and Kahneman (1985)). We argue that the MPL format has the potential to introduce framing effects because the two options are often presented in the same order both within MPL (e.g. with the value difference Δ_t always ascending) and within each binary choice (e.g. with the same option always shown on the left).

The potential for framing bias in binary evaluation tasks is well known; a famous example is the Galaxy Zoo project which asked "volunteer scientists" to assess the rotational direction of images of galaxies. Inflated counts of counter-clockwise rotation in early data led the researchers among other things to vary the arrangement of classification buttons on the screen.[14] Anchoring effects could also arise because of the order within MPL. For example, in an MPL with decreasing Δ_t, the subject may select option A for the first few choices t, based on the utility difference from the options. As she is presented with repeated similar choices, her attention may decrease, or her early choices may even (temporarily) affect her subjective preferences. In both cases she will choose A more often as Δ_t decreases than she would have if presented with the same choice in isolation. In this case, the error term is negative across t (and possibly increasing in size), and as a result, the researcher may overestimate WTP for option A.[15]

Some forms of framing and anchoring in MPL are discussed in the existing literature, e.g. Andersen et al. (2007) (see also section 3.4), but to our knowledge, there is no widespread acknowledgement of potential order biases. Of the 23 studies in Appendix Table B, only three randomize the order of choices within the MPL, and none randomize the order of options within choice.

2.3 Approaches in the literature to MSB, NSB and interval data

Before discussing our proposal for dealing with choice error in MPL, we review the solutions that have been applied in the literature and point out some of their drawbacks. The existing research has introduced a variety of MPL design features aimed at reducing multiple switching and dominated choices *ex ante*, at the elicitation stage. In addition, a range of approaches aim to deal with observed MSB and NSB and the related problem of

---

[14]Between 2007 and 2019, over 400,000 volunteers on Galaxy Zoo carried out more than 11 million classification tasks (Raddick et al., 2019). In early data, it was found that a significantly higher than 50:50 share (around 52%) of galaxies were classified as rotating counter-clockwise. In a test, volunteers were therefore asked to rate mirror images of the same galaxies, and they again classified over 51% of galaxies as rotating counter-clockwise (Land et al., 2008), suggesting that the universe has no directional preference, but humans either have such a preference or click buttons in certain screen locations more frequently.

[15]We assume here that the interest is in a reproducible measure of WTP in contexts outside the MPL.

Table 1: Solutions in the literature for addressing MSB, NSB and interval data.

Stage	Challenge	Solution in the literature	Main issues with the solution
Elicitation	MSB	Elicit a single switchpoint: sMPL, iMPL	Error misattributed to WTP variation, changed outcome variance
	MSB	Stop MPL questions after first switchpoint	Measurement bias; error misattributed to WTP variation, changed outcome variance
Analysis	Inconsistent choices	Drop observations	Selection bias; loss of information, reduced outcome variance
	NSB	Impose minimal/maximal WTP	Measurement bias; potentially changed outcome variance
	Interval data, NSB	Assign single point in the interval, impose min./max. WTP (NSB)	Measurement bias; reduced outcome variance
	Interval data, MSB	Impose interval bounds (MSB) and use interval regression	Loss of information, potentially changed outcome variance

Notes: Overview of common approaches to addressing MPL data challenges arising from inconsistent choices (MSB and NSB with dominated choices) and interval identification (NSB and interval data). "Stage" refers to whether the solution is implemented during data collection (elicitation) or after (analysis). "Solution" and "Main issues" are discussed in more detail in the text.

interval identification when analyzing MPL data *ex post*, at the analysis stage. However, we argue that these solutions do not full address the issues with both idiosyncratic and systematic choice error in MPL. They often misattribute the individual error to variation in WTP, fail to account for order biases, and potentially even introduce new sources of bias, for example through selection. Table 1 summarizes these issues. In this section, we discuss how these solutions can lead to biased WTP measurement or incorrect standard errors. In Section 3.6, we illustrate this using data from South Africa.

At the elicitation stage, MPL design approaches often focus on avoiding or suppressing MSB by enforcing a single switch. The problem is that these solutions may not change the incidence of error, only its visibility. As a result, errors are never observed; variation that may simply be due to error is instead attributed to heterogeneity in WTP. For example, a "switching MPL" or sMPL shows subjects the full MPL and ask them at which point they would switch from option A to option B (e.g., Tanaka et al. (2010), Andersen et al.

(2007)). Andersen et al. (2008) proposed the so-called "iterative MPL" or *iMPL* which narrows down WTP in steps. In Berkouwer and Dean (2021), for instance, subjects are given a price in each step and asked if they would like to buy an improved cook stove. If they say yes, the price is increased, and otherwise it is decreased. In this manner, the procedure arrives iteratively at a small interval for the WTP. An issue with this approach may be that a wrong decision early on artificially increases measurement error.

Another common, and particularly problematic, approach is to elicit pairwise choices only up to the point of the first switch from one option to the other (e.g., Maffioli et al., 2020). Any choice errors in this format will bias WTP downward in price lists with ascending value difference, and upward when the value difference is descending. What is more, if subjects know that the MPL will end as soon as they switch options, it creates an additional incentive to switch early (e.g., if the subject's opportunity cost of time is greater than the utility difference between the options). Note that the bias arising from recording only the first switch works in the opposite direction of any order bias from framing or anchoring described above. This may obscure order effects in data where only the first switch is observed.

Approaches to minimizing MSB that address the underlying sources of error, such as inattention or incomprehension, are rare. In one exception, Yu et al. (2021) successfully reduce MSB by "nudging" subjects to review and potentially revise their answers. Guiteras and Jack (2018) employ a similar approach in eliciting WTA for a piece rate casual labor contract. By repeating each choice four times, and providing an interpretation of the expected take-home pay after each, MSB was effectively eliminated. Note that, by reducing choice error, these approaches may also reduce framing and anchoring biases. However, they also increase implementation costs and may create other biases of their own, such as social desirability bias.

In cases where the data collection procedure does allow for inconsistent choices, methods for dealing with them at the data cleaning or analysis stage tend to introduce new sources of measurement error and bias. Some authors remove inconsistent choices (observations with MSB and sometimes NSB, typically when the experiment includes dominated choices, e.g., Dave et al. (2010)). This approach shrinks the sample size, reduces the outcome variation, and may introduce selection bias.[16]

A challenge presented by all MPL elicitation is the estimation of WTP from the discrete data that arises from a series of binary choices. The experimenter must make

---

[16]Some authors also count the number of times an option was chosen, without regard for the identity of the subjects who made these choices (see e.g. Holt and Laury, 2002).

assumptions about "true" WTP from the identified interval associated with a switch from one option to the other, by selecting either a single location or imposing a distribution of values within this interval. In many cases, either the midpoint or one end point of the interval is used, which results in artificially low variation and may introduce measurement error. In addition, observations that exhibit NSB must either be dropped from the dataset or modified to impose an arbitrary endpoint on the open interval.

The method that most closely respects the structure of the data is the interval regression approach proposed by Andersen et al. (2007), which employs a generalization of the tobit model. However, MSB observations remain a problem; the researcher has to make assumptions about the interval in which the subject's WTP lies for MSB. Typically, the first and last observed switch are used, ignoring information from in-between switches.

In summary, existing approaches to MPL deal with choice error only incompletely. Moreover, the problem of order bias has been largely overlooked. This may be because in many contexts, uniform bias is not overly distorting: for example, in many applications of MPL, the main interest may be ordinal preferences or a simple sample split (e.g., into a more or less risk averse group). However, in WTP elicitation, the (cardinal) monetary value has meaning: a downward or upward shift of the distribution of elicited WTP due to bias may make the difference between a majority of the population expressing a positive vs. negative valuation for a service or policy.

In the next section, we propose a two-pronged approach to MPL design and estimation that uses randomization to detect order biases and a random utility model that includes fixed effects and order indicators to explicitly account for choice error and bias. In section 3.6 we use our approach to show with data from South Africa that (i) there is significant order bias in MPL data that affects estimated WTP, and (ii) many existing estimation approaches may exacerbate measurement issues.

3   Accounting for choice error and bias in MPL design and WTP estimation

3.1   Accounting for bias: MPL design

As discussed above, while the MPL format has many advantages, such as ease of understanding and rich data, it is plausible that it also induces some framing or anchoring effects. A challenge in examining order bias is that typical MPL implementations do not

have enough variation. Specifically, most MPL vary neither the order of binary choices presented, nor the order of options within the binary choice. If option A, say, appears before option B in every binary choice t, then any framing bias towards the first over the second option will be attributed to a higher WTP for option A. Similarly, if subjects are biased towards the option that is associated with the higher value in the first choice (due to anchoring), and all MPL are presented with a descending value difference Δ_t, then this leads to an upwardly biased estimate of the WTP for option A.

The MPL design we propose randomizes both the order of the choice sequence (ascending vs. descending value difference Δ_t) and which option is presented first (e.g. on the left or right, or as the first vs. second option in a verbal question). We call this randomizing *order within MPL*, and randomizing *order within binary choice*.

Randomizing order within MPL allows us to diagnose bias that arises from anchoring effects based on which option has higher value in the first binary choice (see estimation approach below). If order effects are symmetric and the randomization is balanced, it also minimizes the expected net bias in the data. Randomizing order within each binary choice has two roles. First, as above, if subjects have a preference for the option presented first vs. second, random variation in which option appears first/second will allow the experimenter to estimate this systematic error component and minimize net bias. Second, this randomization approach reduces concerns that inattention, or incomprehension, possibly combined with anchoring, could create "never-switching" that is interpreted as a strong preference but in reality only reflects how the MPL is presented.

Specifically, suppose a subject's choices are primarily guided by the layout of the binary choice, say, they always choose the option on the left. In an MPL where all binary choices use the same order, this leads to NSB. In an MPL that randomizes within binary choice, it leads to MSB and the choice pattern is thus correctly attributed to error. In other words, randomization helps distinguish subjects who exhibit NSB simply because they do the exact same thing in each binary choice from subjects who actively choose the same option A or B each time. In addition, it will minimize average bias arising from subjects who always or often choose one side due to reasons unrelated to true WTP.

How does randomization affect the rate of error? As Andersen et al. (2007) have argued, full randomization of binary choices within MPL (i.e. the order of the rows) is unattractive because it makes the MPL task significantly more difficult from a cognitive perspective. We think of randomly presenting choices in ascending vs. descending order as a good compromise. However, we do randomize the order of options within each binary choice. The reader may worry that this may similarly increase attentional cost and lead

to higher error (e.g. the subject might mark the wrong option even after evaluating the options correctly due to the change in side). The randomization within binary choice may also make comprehension harder if the exact parallel organization of each decision helps subjects understand the structure of the experiment. We believe that these risks are small compared to the potential benefits. In fact, we conjecture that within-choice randomization may increase attention, since the choices look less repetitive and are less susceptible to "automation". That said, future work could test the optimal presentation of binary choices to minimize a potential trade-off between detecting hidden, existing error and introducing additional error. Our MPL template survey instrument (described below) allows the user to implement any pattern of randomizing the order of options within binary choice, including no randomization and full randomization.[17]

3.2   Random effects latent utility estimation

Even if the MPL design would allow it, existing estimation approaches do not account properly for either systematic or idiosyncratic error. Building on our framework from section 2.2, we propose to estimate WTP with a latent utility binary choice model that captures both individual-level variation and choice-specific error and can also estimate the size of the systematic bias using data from a randomized MPL. This estimation approach takes seriously the structure of the data as a series of choices between two options, and it deals in natural ways with MSB, NSB, and interval-identified observations. The basic idea appeared early on in the literature on estimating risk aversion parameters from MPL (Holt and Laury, 2002; Harrison and Rutström, 2008; Harrison, 2008), and it is partially reflected in the use of interval regression for MPL data as well as in random-utility approaches found in the wider literature on WTP elicitation (e.g. Berry et al., 2020).[18] However, binary choice models do not seem to be commonly used to estimate preferences from survey-based MPL experiments.

For simplicity, let us assume we have MPL data that compares only two options A

---

[17]In line with the view that full randomization is not desirable, the template does not support a fully random order of choices within MPL. This would require additional coding.

[18]Recently, (Apesteguia and Ballester, 2018) pointed out that the estimation of time and risk preferences from MPL data with a random-utility model may be affected by the fact that the (cardinal) utility difference between two options may not be monotonic in the risk or time preference parameter. As Conte and Hey (2018) note, this non-monotonicity is a property of the (systematic component of) the utility function; issues may arise here from pooling subjects and implicitly making comparisons across subjects. In our case, the monetary value serves a "normalizing" function. We essentially equate WTP with the (relative) subjective value of each option to the subject, and so the utility difference is, by definition, monotonic in WTP.

and B. The dependent variable is an indicator y_{it} that equals 0 if option B is chosen and 1 if option A is chosen. Following Section 2.2, we assume that person i chooses option A, and thus $y_{it} = 1$, if

$$WTP_i + v_{At} - v_{Bt} > \epsilon_{iBmt} - \epsilon_{iAmt}.$$

Here, the suffix m simply denotes the different ways in which the MPL may be presented. We substitute $\epsilon_{iAmt} - \epsilon_{iBmt} = -e_{it} + b_{mt} + s_{mt}$, where b_{mt} and s_{mt} capture order biases in the error term that depends on m and t (see below), and the remaining term $-e_{it}$ is assumed to have mean zero (the negative sign is assumed for convenience). We also decompose WTP_i into average willingness to pay WTP and an individual-specific component η_i. Finally, as above we let $\Delta_t = v_{At} - v_{Bt}$x. Therefore, option A is chosen if

$$WTP + \eta_i + \Delta_t + b_{mt} + s_{mt} > e_{it}.$$

We adopt the convention that the binary choice t is synonymous with the value difference between options Δ_t and that Δ_t is decreasing in t. The MPL may be shown to subjects in ascending or descending order of Δ_t. If the MPL is shown starting at $t = 1$, Δ_t is descending and option A is most attractive relative to B early in the MPL. If the presented choices start at $t = T$, the value difference is ascending, and B starts out most attractive and then becomes less so in the course of the MPL. The order of choices is captured by the index m. In addition, especially in visual MPL presentations, different MPL versions m may reverse the order in which option A and B are presented to the subject within a given binary choice t (i.e. the MPL may vary side of screen or side of page).

We capture any effects that the variation in order has on choices with two bias parameters, b_{mt} for the order of choices within MPL, and s_{mt} for the order of options within binary choice. Suppose for some subject i, the value Δ_t starts high and declines, and the subject chooses A in the first presented choice. As discussed in Section 3.1, the subject may continue to choose A longer than if A started with a lower value relative to B. This situation would be captured by a "boost" to option A relative to B, and therefore a positive b_{mt}, and conversely for cases where the value difference is ascending. Similarly, s_{mt} captures whether one option is favored due to ordering within the binary choice. For example, if subjects prefer options shown on the left of the MPL, we would expect s_{mt} to be positive when A is on the left and negative when it appears on the right.

Assuming a normal distribution for the error term, we can estimate this model with a random effects probit procedure. To do so we write the probability of choosing option A

18

as

$$P(e_{it} < \alpha + \Delta_t + \beta_b x^b_{mt} + \beta_s x^s_{mt} + u_i),$$

where x^b_{mt} and x^s_{mt} are appropriately defined vectors of dummies used to estimate order biases (see below), the constant α provides an estimate of the average willingness to pay, and the term u_i accounts for individual variation in WTP. Even after controlling for preference variation and order biases, the e_{it} may be correlated within MPL, and we recommend clustering standard errors at the subject level.[19]

3.3 Identification of the WTP probit model

Average willingness to pay α. In a probit, utility is taken to be ordinal and all terms are scaled so that the variance of the error term equals 1. However, here we would like to express all terms, and in particular willingness to pay α, in terms of money. We achieve this by restricting the coefficient on Δ_t to be 1 and letting the error variance σ_e be identified off of the data. Equivalently, we can estimate a (standard normal) probit model without the coefficient restriction and re-scale all coefficient estimates. The coefficient estimate for Δ_t represents the inverse of σ_e in this case.

Individual willingness to pay $\alpha + u_i$ and choice error e_{it}. In most latent utility models, the error term is assumed to capture unobserved variation in true underlying utility. Here, we differentiate between the individual-level willingness to pay $\alpha + u_i$, and error e_{it} due to inattention, lack of comprehension, or near indifference (see section 2.2). The two distributions are separately identified because we observe a panel of T choices for each subject, as long as preferences are stable during the administration of the MPL, so that inconsistencies in expressed preferences within the same MPL are the result of choice errors. Both terms are assumed here to have a normal distribution, but other distributions are in principle possible; the most obvious being a logistic distribution for e_{it} which leads to a random effects logit.

The distributional assumptions on u_i and e_{it} discipline the distribution of WTP within each WTP interval. In the probit, the assumed distribution of latent indifference points across MPL intervals is normal with mean α. The location of the sample average will thus partly determine where the mass of indifference points within a given MPL interval is placed. If this seems restrictive, note that any method deriving point estimates for WTP from MPL data must make distributional assumptions. For example, when assigning the

---

[19]For datasets with multiple MPL per subject, we recommend clustering at the subject-by-MPL level.

midpoint of the interval as the switchpoint, the justification is often that any value in the interval is a priori equally likely and that the midpoint represents the average of these values. However, assigning the midpoint ignores the uncertainty from the distribution of WTP within the interval. Moreover, at the population level, this argument implies a "step function" for the probability distribution of WTP that is sensitive to the choice of MPL cutoff points and for which no latent utility model can provide a consistent explanation.

A drawback of the random-effects probit approach is that the model cannot distinguish individual variation in preferences from choice error when the subject does not exhibit MSB. For example, suppose the error terms e_{it} are strongly correlated within subject, perhaps due to anchoring. We cannot distinguish if a subject always chooses B due to anchoring or due to an idiosyncratic, strong preference for option B. This highlights the problem that even choices that are internally consistent may contain undetected error. In the most extreme case, no subject may exhibit MSB, which means that the estimated variance of the error σ_e will tend to zero, and the random-effects probit model may not be identified. Alternative estimation approaches in this situation are a probit without random effects, or interval regression (tobit). Both retain the normal distribution assumption but do not distinguish between systematic preferences and choice error. We argue that the absence of MSB does not guarantee the absence of error; subjects may grasp the structure of the MPL and therefore make choices consistent with a single switchpoint, yet still make errors when indicating this switchpoint. As discussed above, randomizing order within binary choice may help to reveal underlying choice error by inducing MSB in such cases.

Order bias b_{mt} and s_{mt}. So far we have not placed any restrictions on the order bias terms. Note, however, that a bias towards option A, e.g. when A appears high value first, measured by some dummy $x_{mt}^{b+} \in \{0,1\}$, and a bias for option B and against option A, measured by another dummy x_{mt}^{b-}, creates a collinear set of variables when only one MPL is available. Intuitively, if we see a higher WTP in a descending MPL than in an ascending one, we do not know if this is due to the order bias in favor of option A, or the order bias in favor of B, or both. Absent other information, a natural assumption is that the order bias symmetrically favors whichever option appears high value first. If we additionally assume that this order effect on preferences is constant across binary choices t, we might define a combined indicator x_m^b which equals 1 when the MPL starts at $t = 1$ (descending Δ_t) and -1 when it starts at $t = T$ (ascending Δ_t). The coefficient on x_m^b measures the perceived monetary "boost" across binary choices for the MPL option that appears high value first. It is also possible to estimate the bias specific to each choice t,

using T variables x_{mt}^b that take values -1 and 1 in ascending vs. descending MPL in choice t and 0 otherwise. In a parallel manner, we may define x_{mt}^s to equal 1 when A is presented on the left and -1 when it is on the right (either across choices or choice specific).

In order to relax the symmetry assumption and pin down the relative size of the bias for each option without additional MPL data (see below), researchers may choose to conduct outside validation. For example, they could ask (some) subjects to review and potentially revise the MPL responses, as in Yu et al. (2021), or measure the rates at which subjects renege on their original decision, as in Maffioli et al. (2020).

3.4 Extensions

Multiple MPL. We assumed above that we are measuring only one WTP parameter which expresses the relative willingness to pay between options A and B. In practice, researchers may want to estimate the WTP of the same subject for different options from different MPL. This can be achieved in one estimation by creating dummy variables that identify WTP for different option pairs. When estimating relative WTP between a set of different options, it is also possible to impose constraints on the WTP coefficients, such as transitivity. For example, the researcher may estimate $P(e_{it} < \alpha^1 z_m^1 + \alpha^2 z_m^2 + \alpha^3 z_m^3 + \Delta_t + \beta_b x_{mt}^b + \beta_s x_{mt}^s + u_i)$, where α^1 denotes the relative WTP for option 1 over option 2, α^2 denotes the relative WTP for option 2 over option 3, α^3 denotes the WTP for 1 over 3, and the dummies z_m^j indicate the corresponding MPL. We would expect that $\alpha^1 + \alpha^2 = \alpha^3$.

Covariates and WTP heterogeneity. The researcher may also want to allow for heterogeneity in WTP for groups of subjects or examine treatment effects of an experimental intervention. This can be straightforwardly accommodated by including group dummies or covariates and treatment indicator variables.

Other framing effects. We have focused our arguments on order biases. However, our proposed approach offers a framework for investigating other sources of variation in measured WTP. For example, as in Andersen et al. (2007), the researcher might want to test the effect of varying the range of values covered – the range of Δ_t – or the total number of binary decisions T within a given range. Differences in elicited WTP can be measured using dummies, in the same way we estimate order effects above. Again, the researcher would have to impose assumptions about the relative size of bias in any two different ways

of framing the MPL, or find a method to externally validate the estimates, for example by comparing WTP to market price, as in Andersen et al. (2007) and Burchardi et al. (2021). Alternatively, the researchers can impose that the bias is symmetric and the true WTP is the average of the estimates from differently framed MPL, or design an experiment that can measure the framing-specific order bias.

Option-specific order bias.   As discussed above, it is possible that framing biases are not symmetric; in the case of order bias this would be true if option A receives a stronger "boost" with descending Δ_t than option B does with ascending Δ_t. With just one MPL, the two bias terms are not separately identified. However, if subjects complete a set of MPL in which multiple options are all compared with each other, an option-specific order bias can be estimated. This would require at least 3 options and 3 MPL that test all option combinations in the same experiment (see also example in "Multiple MPL" above). We briefly discuss an example of this in section 3.6.

3.5   Implementation

The baseline specification as well as all extensions can be implemented in the Stata procedure included in our estimation package ((Jack et al., 2022), see section 4 and Appendix S4). The user provides a data set with MPL, individual, and choice IDs, the binary outcome of each choice, and a variable containing Δ_t. In addition, covariates, group dummies, or order and framing dummies can be specified. The `mplwtp.ado` file uses the pre-programmed `xtprobit` routine for panel data and rescales the coefficients and standard errors so that the coefficient on Δ_t equals 1. The user can carry out various diagnostics and choose between scaled point estimates and standard errors generated by the delta method or a cluster bootstrap.

3.6   Application: WTP for prepaid electricity credit in South Africa

We use a field experiment in South Africa to demonstrate our approach to measurement and estimation and show evidence that order effects cause bias in MPL measurement. As described in greater detail in Jack et al. (2022), the project used MPL to measure WTP for prepaid electricity (delivered as a voucher that could be directly loaded onto the meter), with the original purpose of understanding the role of transaction costs and liquidity constraints.

Here we use the experimental data to illustrate four things: (1) implementation of the randomized order within MPL and order within binary choice, (2) MSB and NSB in these data, (3) estimation with a scaled random-effects probit, and (4) alternative approaches in the literature to addressing inconsistent choices and interval identification.

Implementation. Three different incentivized MPLs were administered as part of a survey and each subject was randomly assigned to receive two out of the three MPLs (N=767).[20] Here, we focus on one of the MPLs, administered to 506 subjects, which elicited a relative preference for receiving a transfer in the form of an electricity voucher or cash. The other two MPL elicited preferences over one voucher vs. receiving two vouchers on two different days, and two vouchers vs. cash. We include results for estimating WTP in all three MPL in Appendix A (see also below). One binary choice within one of the MPLs was drawn for implementation and the payoffs were immediately realized. We discuss some of our implementation choices in more detail in section 4.

The MPL randomized order both within MPL and within choice. We developed a SurveyCTO template and accompanying randomization files that can be used to carry out an MPL with these features and are described in section 4 and in Appendix S2. The survey module uses a visual representation of the options and shows binary choices one at a time. Subjects select their choice by tapping the screen of the tablet (with the help of an enumerator, if needed).

The structure of the MPL – here with descending value difference Δ_t and showing options in each choice in the same order – is shown in Figure 2. The MPL is designed so that the highest possible value that can be obtained in each choice was 100 Rand. By varying the value of the other option, the value difference increases strictly monotonically between binary choices.

MSB and NSB in our data. Figure 3 shows the number of "switches" observed in the MPL. Around 30% of subjects always choose electricity or always choose cash. Another 46% of subjects switch exactly once. The remainder – 24% – exhibit MSB.[21] It is apparent that there is a higher share of odd than even switches, suggesting that subjects sometimes switch by mistake and then correct themselves. For example, 8% percent of subjects (33%

---

[20]Subjects also went through one practice MPL before being presented with real choices.

[21]We code 12 subjects who switch from the option whose value is ascending to the option whose value is descending as switching twice (thus counting the first choice of the option with increasing value as a "switch"). This avoids lumping inconsistent and consistent choices.

Figure 2: MPL used to measure WTP for prepaid electricity in South Africa

	Electricity vs. cash	
Binary choice	Electricity voucher value	Cash value
t	Option A	Option B
Choice 1	R100	R60
Choice 2	R100	R80
Choice 3	R100	R92
Choice 4	R100	R97
Choice 5	R100	R99
Choice 6	R99	R100
Choice 7	R97	R100
Choice 8	R92	R100
Choice 9	R80	R100
Choice 10	R60	R100

Notes: choices 1-10 were randomly shown in order (decreasing value difference) or in reverse order (increasing value difference). In addition, option A (electricity) and B (cash) were randomly shown either on the left or the right of the screen. Each binary choice was shown separately, accompanied by images showing the amounts either on a cash bill or on a stylized voucher. The respondent made choices by tapping directly on the image on the screen and then confirming the selection. See also figure S1 in appendix S2.

of those exhibiting MSB) have three switches, consistent with one error and subsequent correction.

The rate of MSB is relatively high, although not completely out of line with other MPL data from low-income populations. As discussed, one reason could be that within-choice randomization reveals errors that may otherwise remain undetected (e.g., in the form of NSB). That said, it is possible that the within-choice randomization increased error due to comprehension issues or difficulty using the screens.

The share of never-switchers in the data is also high. Since options were randomized within binary choice, it is unlikely that this is due to inattention that made the subject simply repeat the same choice many times. Only one individual always chose the same side of the screen, that is, exhibited a form of framing-based never-switching. The never-switchers give us a first opportunity to look into the magnitude of order biases. Never switchers are an interesting case because their choices are consistent with a very strong preference for one of the two options. Figure 4 shows the proportion of never-switchers who prefer the electricity voucher.

The left panel shows NSB subjects as a whole. In our data, 54% of all never-switchers

Figure 3: NSB and MSB in South Africa data
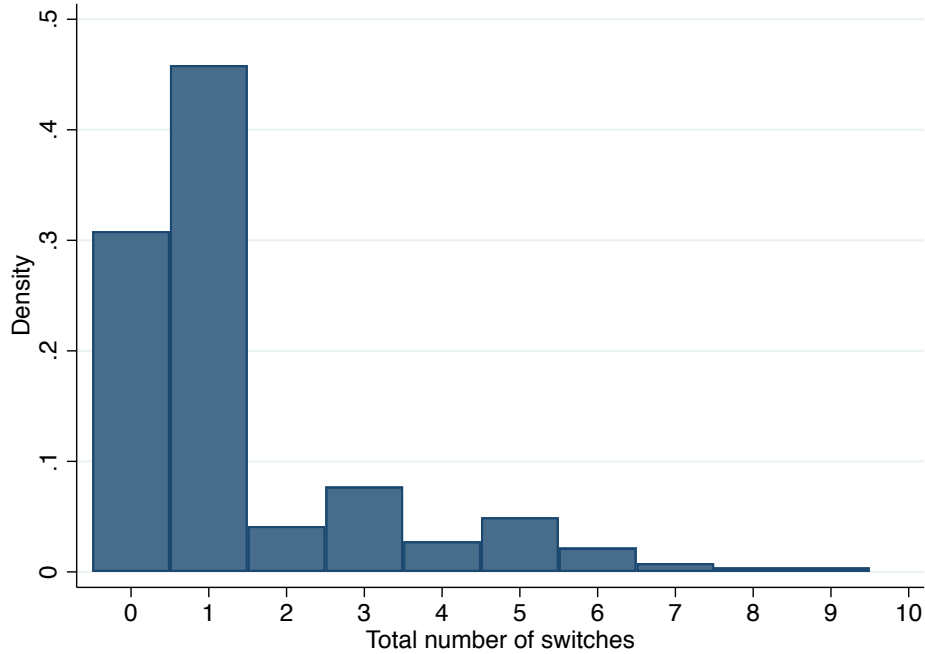
Notes: Total number of times the subject switched between options in the 10-item MPL. 0 switches indicate NSB. 0 or 1 switch are consistent with utility maximizing behavior. 2 or more switches constitute MSB and indicate inconsistent choices.

express a strong preference for the voucher, while 46% choose cash in all choices. Thus, on average, this group has a slight preference for electricity. However, the right panel reveals that the share of NSB who prefer the voucher varies depending on the order in which the choices were presented. The figure shows the share of never switchers who always choose the voucher conditional on whether the value of the voucher was descending or ascending in value. When the value is descending, the share of NSB choosing electricity is 63%. However, when it is descending, the share is only 45%. At the sample level, this suggests that order biases can be quite powerful and shift preferences towards the high-value option in the first MPL choice. At the individual level, however, we cannot distinguish errors or order biases from true preferences. In Jack et al. (2022), we consider reasons why respondents may have strong preferences for both vouchers or cash: frictions due to liquidity issues and transaction costs, combined with unexpected shocks, may lead households to experience unplanned shortages and exhibit a (potentially temporary) high marginal rate of substitution of one for the other.

Figure 4: Share of subjects always choosing voucher, among all NSB
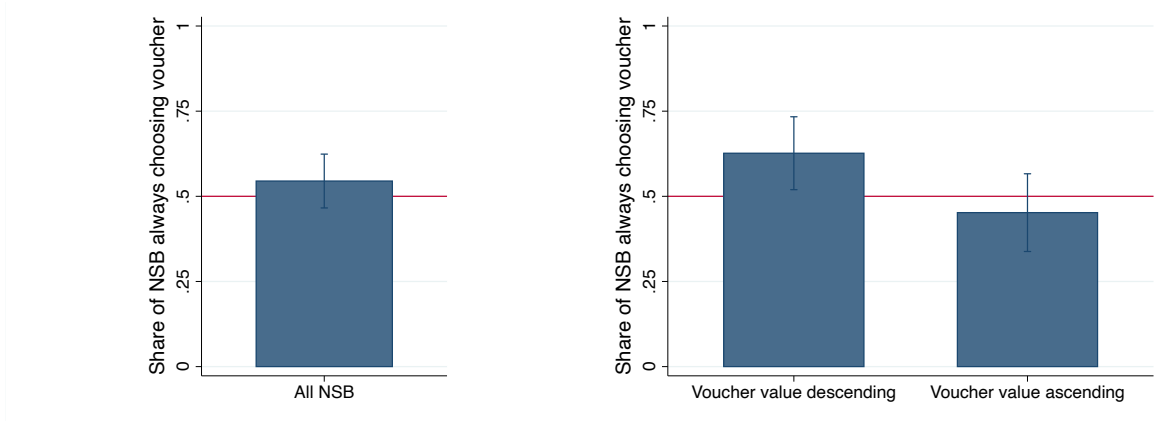
Notes: The figure includes only subjects who exhibit never-switching behavior (NSB), that is, they chose either always the electricity voucher, or always cash. The left panel shows the total share of NSB subjects choosing the voucher always. The right panel shows the share choosing the voucher always, conditional on the order in which the MPL was presented (voucher value descending or ascending). The red line marks an even share of 0.5.

Latent utility estimation. Next, we show results from implementing the random effects probit estimation as in Section 3.2. As described, we scale the estimated coefficients to express all terms in money (here South African Rand).[22]

The first row of Table 2 shows estimated WTP for receiving an electricity transfer over cash in four different model specifications. Columns (1) and (2) restrict the sample to either decreasing or increasing value difference, to mimic many standard MPL designs. In column 1, the value difference is decreasing, so option A (electricity) is higher relative value at the start; in column (2) it is increasing. Consistent with the anchoring effects we discussed above, the WTP estimate for a voucher over cash is positive and significantly different from zero in the first column, but negative and imprecisely estimated in the second column. Using one or the other MPL design, we would have drawn very different conclusions about the value that households attach to receiving a voucher over cash. This is evidence for the importance of randomizing order: it shows that a single ascending or descending MPL would have resulted in biased WTP estimates, even when using the latent-utility estimation approach.

In column (3), the data are pooled so that approximately half of the sample are

---

[22]In Table 2, we show results with standard errors calculated using the delta method. In Appendix table A.1, we show results after bootstrapping the estimation. See Appendix S4 for further discussion of these approaches.

Table 2: Scaled random effects probit estimates.

	Biased sample		Full sample	
	(1)	(2)	(3)	(4)
WTP for elect. voucher	8.879***	-2.360	4.015**	3.590*
	(2.58)	(3.13)	(2.01)	(2.00)
Order within MPL				5.811***
				(2.00)
Order within choice				0.435
				(0.65)
SD of individual error	20.950	26.339	23.388	23.377
SD of RE	39.551	44.458	42.111	41.676
N(choices)	2,710	2,350	5,060	5,060
N(subjects)	271	235	506	506

Notes: Columns 1 and 2 use observations from subjects who saw the value of the electricity voucher (option A) descending and ascending, respectively. Columns 3 and 4 pool all observations. Column 4 adds order controls. Standard errors are clustered at the subject (MPL) level.

presented with either order within MPL. Note that an MPL design with descending Δ_t as in column (1) would have led us to conclude that subjects are willing to pay a nearly 10% tax on a transfer in order to receive that transfer in the form of an electricity voucher. Columns (3) demonstrates that this yields a WTP for the voucher that is 100% higher than the sample average.

The table also reports the standard deviation of the choice-specific error term and the random effect. Both are large, indicating considerable variation in preferences as well as significant error rates – a reflection of the high share of NSB and MSB in our data.

In Appendix A, Table A.2, we show estimates for all three of the MPLs that we implemented in the original data collection (pairwise comparisons of one voucher immediately, two vouchers sent two days apart, and cash). We note that the individual error variance is somewhat lower, and the WTP for one voucher is higher and highly significant, in the MPL that compares one electricity voucher and two electricity vouchers. Both options are in the same "domain" and it appears that subjects had stronger preferences; it is possible that outside factors or inattention therefore confounded elicited preferences less.

Appendix Table A.2, column (4) also demonstrates that we can pool the data and

estimate relative willingness to pay for each pairing of options offered at once. This approach makes it possible, for example, to test joint hypotheses about the coefficients, such as transitivity.[23]

Order biases. Column (4) of table 2 adds basic controls for order biases, by defining a variable that equals 1 (-1) when the MPL is descending (ascending), and a second variable that is 1 (-1) when option A is on the left (right) of the screen. The coefficients on these variables measure the average bias for the option shown with high value first ("order within MPL") and on the left ("order within choice"), respectively.

The estimates show that order within MPL is an important determinant of choice. Specifically, subjects on average express a preference for the option that appears of higher value first that is equivalent to a payment of nearly 6 Rand. By comparison, order within choice does not seem to affect expressed preferences significantly.[24] Similarly, the estimates in Appendix Table A.2 all show significant effects of order within MPL, but not order within binary choice.

The approach in column (4) of Table 2 assumes both that the bias induced by the order within MPL is constant across t, and that it symmetrically favors option A and option B. An inspection of the proportion of subjects choosing option A for each binary choice by MPL order shows that this is likely a simplification and the bias is not constant for all binary choices. In Appendix A, Table A.2, column (5) we therefore include bias terms for each binary choice t separately. Interestingly, the relative bias first increases, then decreases in t. This suggests that order bias might be more pronounced when the options are closer in value and therefore harder to distinguish. However, the WTP estimates themselves are not much affected by including bias terms for each binary choice.

To investigate the symmetry assumption, column (6) in Table A.2 replaces the symmetric "order within MPL" variable with three variables that separately measure order bias in favor of each of the three options. Note that this replaces the symmetry restriction with the assumption that the bias in favor of a given option is constant across MPL comparing different options. The results suggest that order bias for cash is stronger than for the two electricity options. While the qualitative conclusions about the WTP do not

---

[23]It would also be possible to directly impose transitivity or other restrictions when estimating these three WTP values in one model. This option is not currently implemented in our Stata routine and we do not show it here.

[24]We do not have enough variation in the total number of times options switched sides, so we cannot test interaction effects, that is for example, whether more or less variation in the order within binary choice affects the magnitude of the within-MPL order bias.

change much, the point estimates shift somewhat. Note, however, that a different interpretation of the data is that the (symmetric) order bias is simply lower in the MPL that compares one and two electricity vouchers vs. the MPLs that compare electricity and cash (see columns (1)-(3) of Table A.2). Note also that this approach to estimating bias is only possible with data from at least 3 MPL, and so in many cases may not be a strategy available to the researcher.

(Re-)Introducing bias through standard procedures In Section 2.3 and Table 1, we described some common approaches to addressing MSB, NSB and the interval nature of MPL data. Recall that all approaches discussed above distill the information from a subject's MPL into an interval or even just a single point; some additionally drop inconsistent information or otherwise restrict the data. Here, we replicate some of these approaches to demonstrate how this influences estimated WTP.

We make our points mostly using interval regression estimates. Interval regression is a generalization of a tobit specification and uses one upper and lower bound for the WTP of each subject to estimate average WTP for an electricity voucher. Therefore the researcher has to assign an interval for WTP to each subject based on the observed MPL choices. This is often done by using the first and last observed "switch". Subjects with MSB are thus assigned a wider interval to accommodate all of their choices, and switches between the interval endpoints are effectively not used.[25] Subjects with NSB can be assigned one-sided (open ended) intervals. They may alternatively be assigned a maximum or minimum of WTP based on the researcher's assessment which values of WTP are plausible. For example, in the South Africa data, a WTP below -100 would indicate that a subject would rather pay cash than receive the R100 electricity voucher.

Table 3 columns (1) to (4) implement these two variants of interval regression (unrestricted max./min. WTP or restriction to ± 100). For all regressions in the table, odd columns use observations with descending value difference, and even columns with ascending value difference, mimicking again typical MPL data. They can therefore be compared with columns (1) and (2) in Table 2, respectively. As in the scaled probit, the WTP estimates are sensitive to choice order, but they are also attenuated toward zero; slightly more so when a maximum and minimum WTP are imposed in columns (3) and (4).[26]

Some MPL implementations also suppress MSB observations, either ex ante or ex

---

[25]This follows the approach in Andersen et al. (2007).

[26]Note that the number of subjects is lower in (1) and (2) because of subjects whose upper and lower bounds are both undefined. This happens if a subject chooses option B (the lower-value option) in $t = 1$ but option A in $t = 10$ (again the lower-value option).

post. Columns (5)-(8) continue to use interval regression to maintain comparability, and implement two versions of these sample restrictions. One common approach is to stop the data collection once a subject switches for the first time. We approximate this by using the WTP interval associated with only the first switch and re-estimating the interval regression. As shown in columns (5) and (6), the point estimates change substantially and standard errors increase. Interestingly, comparing column (5) with column (1) shows that the measurement bias from using the first switch essentially erases the order bias in the opposite direction (see discussion in section 2.3). We note that Channa et al. (2021) and Fuller and Ricker-Gilbert (2021) both test WTP for different kinds of grain quality verification. Both papers randomize the order within MPL (a small minority among MPL implementations), but also only use the first switch. Fuller and Ricker-Gilbert (2021) do not find an order effect, whereas Channa et al. (2021) document a strong order effect among farmers but not traders. It is possible that a full MPL elicitation (i.e., not stopping after the first switch) would have shown order bias in all samples. Given the popularity of the "first switch" MPL, it is possible that other researchers have piloted order randomization before us but dismissed order bias concerns in MPL based on the data.

Another, fairly drastic approach to addressing inconsistency is to throw out observations with MSB. We implement this in columns (7) and (8) of Table 3. In our data and within the interval regression approach, excluding MSB leads to more extreme estimated WTP (depending on order within MPL). We are agnostic whether this is due to selection bias or a sharper estimate, but note that the reduced sample size shows that we would be discarding a large number of data points. This seems in general an unsatisfactory approach.

Many times, MPL data analysis does not use interval regression, but instead reduces the WTP interval to a single point and assigns the minimum, midpoint or maximum of the interval as the individual measure of WTP. In our data, this has a drastic effect on implied WTP. For example, when imposing a WTP minimum of -100 and maximum of 100 (for NSB observations) and using the minimum, maximum or midpoint of the WTP intervals defined by the first and last switch, we get an average WTP of -18.0, 2.7 and 23.4 Rand, respectively.

Overall, neither the estimates from Table 3 nor the single-point estimates reliably reproduce our scaled-probit results, and the different methods lead to a dizzying array of estimated WTP levels. This confirms our conjecture from section 2.3 that the way the literature has dealt with choice error is inconsistent and unreliable in the context of

Table 3: Alternative approaches: Sample or data restrictions

Sample/Data	Inf. interval		100 endpoint		First switch only		No MSB	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WTP	6.366***	-1.551	6.086***	-1.459	1.824	3.695	8.732***	-3.943
	(2.45)	(2.84)	(2.31)	(2.62)	(2.73)	(3.11)	(2.87)	(3.96)
N(subjects)	266	228	271	235	271	228	223	166

Notes: Interval regression estimates, using the first and last switch to define the interval in the case of MSB. Columns 1 and 2 allow an infinite interval for NSB. Columns 3 and 4 impose an endpoint of ± 100 on the open interval for NSB observations. Columns 5 and 6 use the first observed switch only. Columns 7 and 8 drop subjects who exhibit MSB. In odd-numbered columns, the value of the voucher is descending; in even-numbered columns, it is ascending.

WTP estimation. In particular, the single-point approaches yield WTP estimates that are so sensitive to the choice of point within the interval to be more or less useless. But even the interval regression estimates deviate from the probit estimates. Moreover, in all cases order effects reverse the sign and alter the significance level of the WTP estimate, emphasizing the need for order randomization.

We note one exception regarding using scaled probit over interval regression, and that is when subjects do not exhibit MSB. In this case, which may be more common in high-income settings or in the laboratory, interval regression or simple probit is the correct approach due to the identification problems in random-effects probit when there are no observed choice errors (see above).

4 MPL implementation

In the course of our own work and while surveying the literature, we identified a number of other design and implementation features besides order randomization that can help to decrease bias and individual error. We discuss some of these design features below. The SurveyCTO template that we provide as part of this paper, described in Appendix S2, implements these features along with a very flexible specification for order randomization. Many of these features may also be beneficial in other implementations of MPL, such as pen-and-paper questionnaires or laboratory experiments.

4.1 Best practices

Keeping maximum payout constant. As seen in Table 2, the MPL we used in South Africa keeps the maximum payout $\max\{v_a, v_b\}$ in each choice constant. This feature is also implemented in Allcott and Kessler (2019), see Figure 1. Keeping overall value constant makes it more likely that attention levels remain similar across the MPL. Moreover, the MPL incentive appears fair to subjects: regardless which binary choice is selected for implementation, the maximal monetary value the subject could have obtained remains the same. In manual implementations of an incentivized MPL where the realized choice is determined by a draw or dice roll, it also avoids irregularities in the randomized selection of the implemented choice.[27]

Value difference decreases. It can be helpful to elicit smaller WTP (value difference) intervals around points that matter for hypothesis testing. In the South Africa Electricity example, this is the point where the value difference and therefore WTP is zero. If the interval where subjects switch from one option to the other is small, the estimation of willingness to pay will be more precise. We recommend calibrating the value steps in the MPL in piloting. Again, this feature is also implemented in Allcott and Kessler (2019).

Binary choices presented separately. If MPL are displayed as a list on a single page, some subjects may evaluate the MPL as a whole instead as a series of binary choices. This may increase anchoring effects. From our piloting work we concluded that showing each new binary choice on a new page or screen, combined with within-choice randomization, helps increase attention.

Practice MPL. Before the MPL of interest is implemented, it is in our experience helpful to let subjects complete a practice MPL. Our practice MPL pays out a randomly selected binary choice for items of relatively low value and unrelated to the research question; in South Africa these were types of candy. In addition to improving comprehension of the format of an MPL, demonstrating the realization of one binary choice increases attention and emphasizes the independence of individual binary choices. We recommend short practice MPLs (2 - 3 binary choices). Their format should be as similar as possible to the actual MPL.

---

[27]One of the authors (Sautmann) found in a pilot implementation of a time-preference MPL that the distribution of randomly drawn implemented choices was markedly skewed towards high-value choices, likely because surveyors let subjects take a second random draw when the first draw yielded a low payout.

As part of the practice MPL, subjects may review the choices they made in each practice binary choice. They may be asked to consider whether they would have been pleased with the MPL payout if a given choice had actually been randomly selected. This can decrease the chance that subjects view the MPL as a test with a "correct" answer. Many papers carry out practice elicitation, including, for example, Fuller and Ricker-Gilbert (2021).

Formatting. An engaging design and clear formatting and layout can increase attention, improve understanding, and decrease fatigue and mistakes. When subjects complete the MPL themselves, formatting choices such as a large font size and images can increase comprehension, including in contexts with low literacy. Attention may also be increased by prompting subjects to verify their response to a binary choice before moving on to the next one. Other visual aids, such as showing an animated coin flip at the end of the MPL, can help convey features of the MPL, here that the realized choice is selected randomly.

4.2 Implementation Package

As part of this paper, we provide an implementation package for MPL data collection, processing, and analysis, described in detail in a technical appendix.

The package can create a SurveyCTO template that implements the design features above, along with randomization within MPL and within binary choice. SurveyCTO is a common survey tool used for conducting in-field surveys via an app on a phone or tablet. The SurveyCTO file is fully automatically generated from a Stata script together with a set of user-defined inputs, as described in Appendix S2. The template supports flexible randomization specifications for various elements of the MPL and can accommodate a pre-existing survey sample. With small changes, the SurveyCTO template can be used for a wide range of preference elicitation experiments, including risk aversion or time preferences. It can also be adapted for other Open Data Kit (ODK) data collection platforms.

We also provide Stata programs that prepare the data generated by the SurveyCTO questionnaire for analysis (Appendix S3), and an ado file that specifies a command to estimate WTP with our proposed scaled random-effects probit (Appendix S4). The command can output standard errors via the delta method or via cluster bootstrap and supports various diagnostics to ensure accurate estimates.

5   Conclusion

Measuring willingness to pay for goods and services is an important part of many research studies. Among other things, information on WTP can provide insights about welfare (e.g., Allcott and Kessler, 2019), allow cost-benefit comparisons (e.g., Cole and Fernando, 2020), and help explain treatment effects (e.g., Guiteras and Jack, 2018). Multiple price list elicitation is attractive for many reasons. It offers a compromise between the precise information revealed by BDM and the simplicity of TIOLI, and data from repeated choices can in principle be used to learn about error from inconsistent and dominated choices. As we show in this paper, in combination with standard MPL design features this error can introduce bias in WTP estimates.

We propose a straightforward, two-part approach to revealing both idiosyncratic and systematic choice error and incorporating it into WTP estimation. First, random variation in the order of choices within MPL and within binary choice reveals bias and minimizes its net impact on the estimated WTP. Second, latent utility estimation using random effects probit accommodates multiple switching, never switching and interval data, and can be used to estimate bias terms and both individual choice error and subject-level variance in WTP. To support other researchers interested in adopting these innovations, we offer a SurveyCTO package and a Stata ado file, along with the necessary instructions.

We focus on a single case study to demonstrate both the challenges with alternative approaches and the strengths of our approach. This case is intended to be illustrative and we are well aware of the limited general conclusions we can draw. For example, WTP estimation in populations that exhibit less error and bias will benefit less from both the randomized choice implementation and the estimation approach. Note, however, that only by randomizing the order of choices in the MPL and options within each binary choice researchers can test for any framing or anchoring effects in new samples.

A second limitation of our approach is that the correction of order bias can be applied to the study sample, or to large enough subgroups within the study sample, but likely not to individual-level data. The implication is that MPL data – carefully implemented and analyzed – is appropriate for measuring sample average WTP, but not individual-specific WTP. It should also be noted that inattention, incomprehension, and framing effects can lead to choice error in other elicitation methods too. Given the prevalence of these errors in WTP estimation, the method we propose in this paper offers a way forward: an estimation approach that correctly accounts for individual error and a measurement approach that reveals bias.

In the MPL implementation in this paper, we only varied whether the MPL has an ascending or descending value difference, while we allowed order within each binary choice to be fully randomized. Future work might test whether the degree of random variation in MPL implementation can be further optimize to minimize the incidence of biases and errors. Specifically, there could be a trade-off between too little randomization introducing net bias and preventing the detection of latent choice errors, and too much randomization increasing the cognitive burden and thereby provoking choice error that did not previously occur.

References

Agranov, M. and P. Ortoleva (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy 125*(1), 40–68.

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–76.

Alphonce, R. and F. Alfnes (2017). Eliciting consumer WTP for food characteristics in a developing context: Application of four valuation methods in an african market. *Journal of Agricultural Economics 68*(1), 123–142.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2007). Valuation using multiple price list formats. *Applied Economics 39*(6), 675–682.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences. *Econometrica 76*(3), 583–618.

Apesteguia, J. and M. A. Ballester (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy 126*(1), 74–106.

Asioli, D., A. Mignani, and F. Alfnes (2021). Quick and easy? Respondent evaluations of the Becker-DeGroot-Marschak and multiple price list valuation mechanisms. *Agribusiness 37*(2), 215–234.

Bartling, B., F. Engl, and R. A. Weber (2015). Game form misconceptions are not necessary for a willingness-to-pay vs. willingness-to-accept gap. *Journal of the Economic Science Association 1*(1), 72–85.

Ben Yishay, A., A. Fraker, R. Guiteras, G. Palloni, N. B. Shah, S. Shirrell, and P. Wang (2017). Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia. *Journal of Environmental Economics and Management 86*, 121–140.

Bénabou, R., A. Falk, L. Henkel, and J. Tirole (2020). Eliciting moral preferences: Theory and experiment. *Working paper*.

Berkouwer, S. B. and J. T. Dean (2021). Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households. *Working paper*.

Berry, J., R. Dizon-Ross, and M. Jagnani (2020). Not playing favorites: An experiment on parental fairness preferences. *NBER Working Paper 26732*.

Berry, J., G. Fischer, and R. Guiteras (2020). Eliciting and utilizing willingness to pay: Evidence from field trials in Northern Ghana. *Journal of Political Economy 128*(4), 1436–1473.

Björkegren, D. (2020). Competition in network industries: Evidence from the rwandan

mobile phone network. *Working paper*.

Burchardi, K., J. De Quidt, S. Gulesci, B. Lerva, and S. Tripodi (2021). Testing willingness to pay elicitation mechanisms in the field: Evidence from Uganda. *Journal of Development Economics 152*, 102701.

Burchardi, K., J. De Quidt, S. Gulesci, and M. Sulaiman (2021). Credit constraints and demand for remedial education: Evidence from Tanzania. *Working paper*.

Bursztyn, L., B. Ferman, S. Fiorin, M. Kanz, and G. Rao (2018). Status goods: Experimental evidence from platinum credit cards. *The Quarterly Journal of Economics 133*(3), 1561–1595.

Cason, T. N. and C. R. Plott (2014). Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. *Journal of Political Economy 122*(6), 1235–1270.

Cassar, A., F. Wordofa, and Y. J. Zhang (2016). Competing for the benefit of offspring eliminates the gender gap in competitiveness. *Proceedings of the National Academy of Sciences 113*(19), 5201–5205.

Channa, H., J. Ricker-Gilbert, H. De Groote, and J. Bauchet (2021). Willingness to pay for a new farm technology given risk preferences: Evidence from an experimental auction in Kenya. *Agricultural Economics 52*(5), 733–748.

Cole, S., A. N. Fernando, D. Stein, and J. Tobacman (2020). Field comparisons of incentive-compatible preference elicitation techniques. *Journal of Economic Behavior & Organization 172*, 33–56.

Cole, S. A. and A. N. Fernando (2020). 'Mobile'izing agricultural advice technology adoption, diffusion and sustainability. *The Economic Journal 131*(633), 192–219.

Conte, A. and J. D. Hey (2018). Rehabilitating the random utility model. A comment on Apesteguia and Ballester (2018). *Working paper*.

Crosetto, P. and A. Filippin (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics 19*(3), 613–641.

Dave, C., C. C. Eckel, C. A. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty 41*(3), 219–243.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies 83*(2), 587–628.

Filippin, A. and P. Crosetto (2016). A reconsideration of gender differences in risk attitudes. *Management Science 62*(11), 3138–3160.

Fuller, A. J. and J. Ricker-Gilbert (2021). Estimating demand for third-party quality testing in rural grain markets: Evidence from an experimental auction for measuring

moisture content in Kenya. *Journal of African Economies 30* (5), 389–417.

Gillen, B., E. Snowberg, and L. Yariv (2019). Experimenting with measurement error: Techniques with applications to the Caltech cohort study. *Journal of Political Economy 127* (4), 1826–1863.

Guiteras, R. P. and B. K. Jack (2018). Productivity in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics 131*, 42–61.

Harrison, G. W. (2008). Maximum likelihood estimation of utility functions using Stata. *Working paper*.

Harrison, G. W. and E. E. Rutström (2008). Risk aversion in the laboratory. In J. C. Cox and G. W. Harrison (Eds.), *Risk Aversion in Experiments*, Volume 12 of *Research in Experimental Economics*, pp. 41–196. Emerald Group Publishing Limited.

Holt, C. and S. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92* (5), 1644–1655.

Jack, B. K., K. McDermott, J. Romine, and A. Sautmann (2022). Paying for power: Utility bills and the spending patterns of the poor. *Working paper*.

Jack, K., K. McDermott, A. Sautmann, K. Chan, and L. Yue (2022). Implementation package: Multiple price lists for willingness to pay elicitation. URL: `https://github.com/MPL-WTP/MPL-WTP`.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy 98* (6), 1325–1348.

Land, K., A. Slosar, C. Lintott, D. Andreescu, S. Bamford, P. Murray, R. Nichol, M. J. Raddick, K. Schawinski, A. Szalay, D. Thomas, and J. Vandenberg (2008, 08). Galaxy Zoo: the large-scale spin statistics of spiral galaxies in the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society 388* (4), 1686–1692.

Lang, G. and B. Lanz (2021). Energy efficiency, information, and the acceptability of rent increases: A survey experiment with tenants. *Energy Economics 95*, 105007.

Maffioli, A., D. McKenzie, and D. Ubfal (2020). Estimating the demand for business training. *Working paper*.

Raddick, M. J., E. E. Prather, and C. S. Wallace (2019). Galaxy zoo: Science content knowledge of citizen scientists. *Public Understanding of Science 28* (6), 636–651.

Squires, M. (2021). Kinship taxation as an impediment to growth: Experimental evidence from Kenyan microenterprises. *Working paper*.

Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review 100* (1), 557–71.

Tversky, A. and D. Kahneman (1985). The framing of decisions and the psychology of choice. In *Behavioral decision making*, pp. 25–41. Springer.

Yu, C. W., Y. J. Zhang, and S. X. Zuo (2021). Multiple switching and data quality in the Multiple Price List. *The Review of Economics and Statistics 103*(1), 136–150.

A Additional results

Table A.1: Bootstrapped scaled random effects probit estimates

	Biased sample		Full sample	
	(1)	(2)	(3)	(4)
WTP for elect. voucher	8.866***	-2.357	4.009**	3.586*
	(2.51)	(3.17)	(1.93)	(1.93)
Order within MPL				5.803***
				(1.96)
Order within choice				0.436
				(0.63)
SD of individual error	20.944	26.335	23.383	23.373
SD of RE	39.555	44.461	42.114	41.679
N(choices)	2,710	2,350	5,060	5,060
N(subjects)	271	235	506	506

Notes: Replication of Table 2, bootstrapping the estimates and standard errors.

Table A.2: WTP estimates for three MPL, separately and jointly

	Separate samples			Full sample		
	(1)	(2)	(3)	(4)	(5)	(6)
WTP for elect. voucher	3.590*			3.416*	3.467*	5.578*
	(2.00)			(1.84)	(1.85)	(3.01)
WTP 1 v. over 2 v.		7.181***		8.556***	8.684***	8.224***
		(1.24)		(1.41)	(1.42)	(3.02)
WTP 2v over cash			-0.209	0.017	0.054	2.548
			(2.10)	(1.95)	(1.96)	(3.01)
Order within choice	0.435	-0.437	0.367	0.071	0.094	0.071
	(0.65)	(0.50)	(0.59)	(0.33)	(0.33)	(0.33)
Order within MPL	5.811***	2.459**	5.267**	4.402***		
	(2.00)	(1.19)	(2.09)	(1.00)		
Order within MPL (1 voucher)						3.140
						(3.02)
Order within MPL (cash)						7.596**
						(3.01)
Order within MPL (2 vouchers)						2.505
						(3.01)
Order within MPL (t=1)					1.506	
					(1.88)	
Order within MPL (t=2)					2.407*	
					(1.44)	
Order within MPL (t=3)					5.509***	
					(1.31)	
Order within MPL (t=4)					5.968***	
					(1.32)	
Order within MPL (t=5)					5.755***	
					(1.31)	
Order within MPL (t=6)					7.024***	
					(1.37)	
Order within MPL (t=7)					7.741***	
					(1.37)	
Order within MPL (t=8)					5.662***	
					(1.33)	
Order within MPL (t=9)					2.557*	
					(1.38)	
Order within MPL (t=10)					0.062	
					(1.68)	
SD of individual error	23.377	20.333	21.336	21.655	21.599	21.655
SD of random effect	41.676	24.009	44.548	36.231	36.513	36.212
N(choices)	5,060	4,990	5,110	15,160	15,160	15,160
N(subjects)	506	499	511	1,516	1,516	1,516

Notes: Scaled random effects probit estimates from three MPLs implemented on a single subject pool in South Africa. All subjects (N=1,516) saw two of the three MPLs. Column 1 reproduces the estimates from column 4 of Table 2. Column 2 shows estimates from MPL comparing one or two electricity vouchers. Column 3 shows estimates from MPL involving two electricity vouchers or cash. Column 4 estimates all three MPLs jointly.

A.2

B   List of papers considered

Table B lists 23 papers that measure WTP using an MPL format, sorted by publication year and within year alphabetically by author. These were discovered and classified in an extensive literature search that served both to understand how the method of MPL has developed and how it is used today. We note that the list is not representative or exhaustive, and due to a snowball search by citations may focus more strongly on recent papers. Part of the challenge was that many papers that elicit WTP do not mention the exact elicitation method they use or describe the approach as BDM, and researchers may use alternative terms for willingness to pay, such as contingent valuation, value, price, etc. On the other hand, a very large number of papers that do mention MPL use it to elicit risk or time preferences. Nonetheless, the large number of papers that measure WTP using MPL from 2020 onward shows that the method is routinely used in current applied research.

The table prioritizes examples from low and middle income countries (LMIC) that were incentivized and carried out "in the field," meaning with a representative sample of the population of interest, often as part of an RCT or larger survey. We do list a few influential papers from high-income countries, often conducted in the laboratory. Several of them were early studies that focus on the method of MPL and compare different elicitation methods or formats.

Most MPL were conducted in lab-in-the-field or laboratory experiments, but we also include three surveys (meaning MPL-style questions were asked but the answers were not incentivized) and two hybrid cases: an experiment carried out on Amazon MTurk with hired participants and a classroom experiment. Many studies elicit WTP for everyday items such as mugs, pens, chocolate or soap.

We also include whether the data collection instrument or data is publicly available. In most recent papers, both are made available at publication. In many of the papers listed, detailed sections in the paper or in an appendix describe the instrument and experimental procedure. Papers in italics were unpublished at the time of writing, so the data may become available at a later date.

Paper	WTP for	Country	Purpose	Data collection	Data/instr. available
Berkouwer et al. 2021	Electricity transfer, electricity or cash loan	Ghana	WTP	Survey	No/Yes
Berkouwer and Dean (2021)	Efficient stove	Kenya	WTP	Lab-in-the-field	No/Yes
Burchardi et al. (2021)	Induced value, everyday items	Uganda	Method	Lab-in-the-field	Yes/Yes
Burchardi et al. (2021)	Remedial education program	Tanzania	WTP	Lab-in-the-field	No/Yes
Channa et al. (2021)	Moisture detection device for maize	Kenya	WTP	Lab-in-the-field	Yes/Yes
Fuller and Ricker-Gilbert (2021)	Third-party grain quality testing	Kenya	WTP	Lab-in-the-field	Yes/Yes
Lang and Lanz (2021)	Efficient heating appliance	Switzerland	WTP	Survey	Yes/Yes
Squires (2021)	Hiding experimental income	Kenya	WTP	Lab-in-the-field	No/Yes
Asioli et al. (2021)	Everyday items	Norway	Method	Laboratory	Yes/No
Bénabou et al. (2020)	Moral action (donation)	Germany	WTP	Laboratory	No/Yes
Björkegren (2020)	Change mobile phone provider	Rwanda	WTA	Survey	No/Yes
Cole and Fernando (2020)	Agricultural extension service by phone	India	WTP	Lab-in-the-field	No/No
Maffioli et al. (2020)	Business training	Jamaica	WTP	Lab-in-the-field	No/Yes
Allcott and Kessler (2019)	Home energy reports	USA	WTP	Lab-in-the-field	Yes/Yes
Bursztyn et al. (2018)	Status good (Armani gift card)	USA	WTP	Amazon MTurk	Yes/Yes
Guiteras and Jack (2018)	Piece rate contract	Malawi	WTA	Lab-in-the-field	No/Yes
Ben Yishay et al. (2017)	Improved latrines	Cambodia	WTP	Lab-in-the-field	No/No
Alphonce and Alfnes (2017)	Organic and conventional tomatoes	Tanzania	Method	Lab-in-the-field	No/Yes
Cassar et al. (2016)	Book voucher	China	WTP	Laboratory	No/Yes
Bartling et al. (2015)	Induced value, everyday items	Switzerland	WTP/ WTA	Laboratory	No/Yes
Exley (2016)	Charitable giving	USA	WTP	Laboratory	No/Yes
Andersen et al. (2007)	Everyday items	USA	Method	Laboratory	No/No
Kahneman et al. (1990)	Induced value, everyday items	USA	WTP/ WTA	Classroom	No/No

Multiple Price Lists for Willingness to Pay Elicitation: Technical Appendix

Kelsey Jack, Kathryn McDermott, Anja Sautmann

Table of contents

S1 Introduction

This appendix describes an implementation package for WTP elicitation available on GitHub at `https://github.com/MPL-WTP/MPL-WTP` (Jack et al., 2022).[1] The package consists of three components described in the three parts of this technical appendix.

In part S2, we describe the Stata scripts and inputs that together are used to construct a SurveyCTO template to carry out MPL-based measurement of willingness to pay. Part S3 describes the data processing used to transform the SurveyCTO data output into an analysis-ready dataset that conforms to the estimation approach described in the paper. In part S4, we introduce a custom Stata command for the data analysis using random effects probit estimation.

The associated folders on GitHub are called "1_Data_Collection", "2_Data_Processing", and "3_Data_Analysis". Folders 1 and 2 are organised into "Code" and "Example" folders. The "Code" folder contains Stata do files that produce the SurveyCTO form and processed data. The "Example" folder contains examples of the inputs needed as well as the resulting outputs.

Throughout this appendix we will use the experiment in South Africa from Jack, McDermott and Sautmann (JMS) as an example. In this experiment we measured willingness to pay between three options: receiving one electricity voucher right away, receiving two electricity vouchers one right away and one at a later point in time, and receiving cash. Every subject was asked two out of a possible three MPLs, each of which offered choices between two of these options.

The different components of the implementation package can be used together or separately. Users who plan to collect original data in SurveyCTO and estimate willingness to pay using those data may use all three components. Other users who have data already in hand may wish to use only the estimation program and refer to the data processing documentation to understand the necessary data set up. Alternatively, some users might collect and prepare the data using the SurveyCTO and processing template files, but choose a different analysis approach.

Even users who only plan to read the data set up and estimation sections of this appendix may benefit from reading the glossary in section S2.1 below. It explains the terms and conventions used throughout.

---

[1] We ask users of any part of the code in this repository to please cite the implementation package in conjunction with this paper.

S2 Data Collection: SurveyCTO Template

This section of the appendix describes the contents and use of a Stata script that creates a multiple price list (MPL) instrument to measure the relative monetary value of two options, such as goods or payments (incl. willingness to pay/willingness to accept). The instrument is implemented in SurveyCTO for use on a touchscreen tablet with a visual interface that can be used by the subject directly or the enumerator and subject together.

This implementation of an MPL shows a series of binary choices one at a time on a screen. Options within each binary choice can be displayed as text or depicted as images so that the MPL can be implemented in settings with low literacy levels. Different components of the MPL instrument can be randomized: the order in which the binary choices are shown (within-MPL variation), the placement of MPL options on the screen (within-choice variation), and/or the order in which multiple MPLs are asked, if more than one MPL is given to one subject). We provide a template for randomizing these components that ensures balance.

The next two sections provide a glossary that defines the terminology that will be used throughout and give an overview over the Stata do files and their output. Section S2.3 describes how the MPL looks in SurveyCTO and how subjects will interact with it. Section S2.4 describes how to define the entries of the MPL. Section S2.5 describes how to set up the Stata do file that creates the MPL instrument. Section S2.6 describes how the randomization is implemented. Section S2.7 describes the SurveyCTO output generated using the Stata do files.

S2.1 Glossary

We use a set of terms to refer to the different elements of the experiment described in the main text. The variable names in the SurveyCTO and Stata templates use the same terms. Figures S1 and S2 show how these terms relate to the survey instrument as it is seen by subjects.

Figure S1: Binary choice example with images.
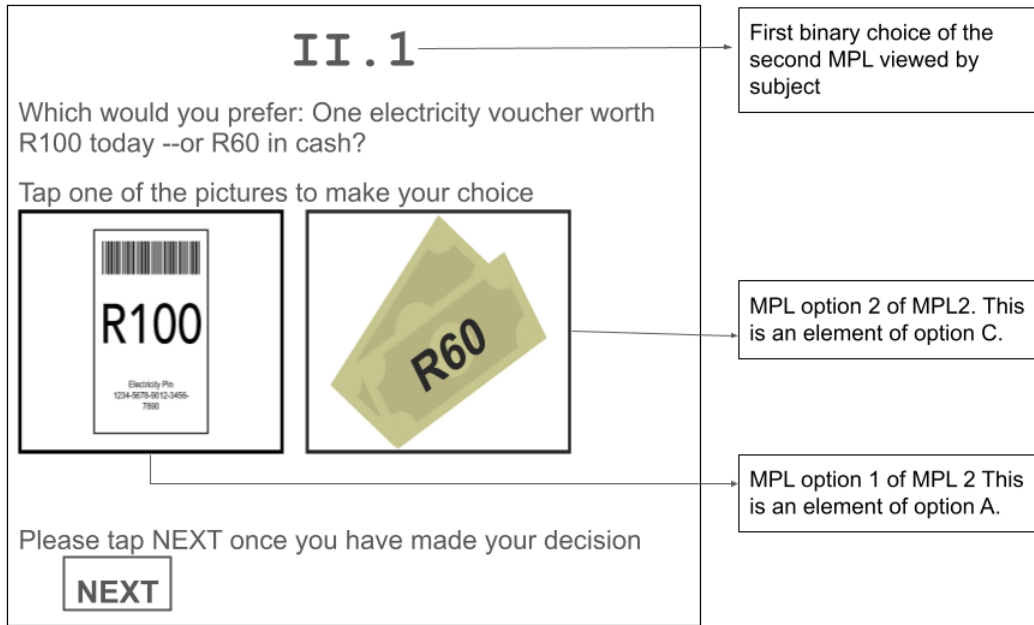
Notes: Binary choices are labelled with roman numerals indicating the MPL order (the picture shows the second MPL done by the subject) and numbers (this is the first binary choice in this MPL). We are showing here an example of MPL2, which compares one voucher to cash. Monetary values are in South African Rand.

Figure S2: Binary choice example with no images
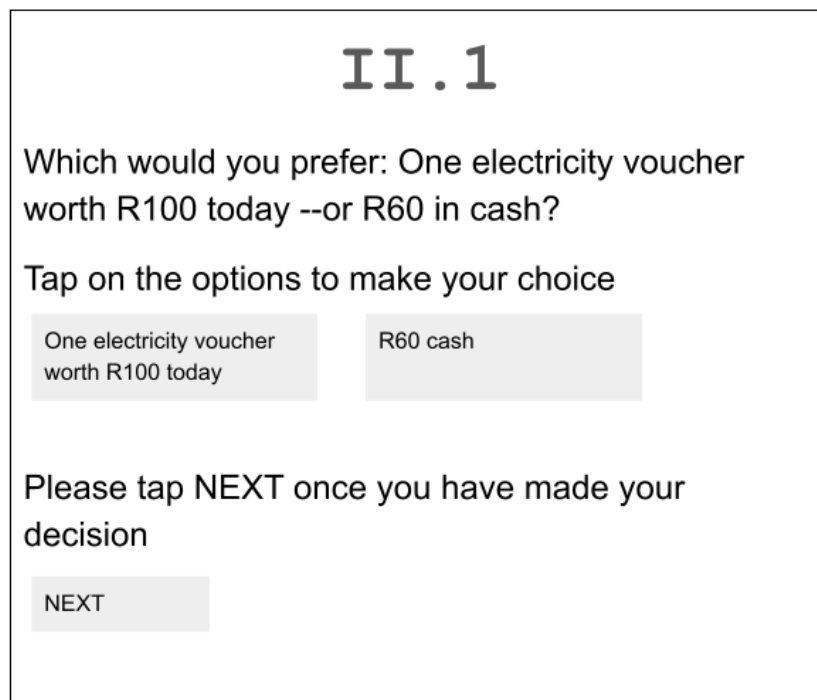
MPL An MPL (Multiple Price List) elicits the subject's relative preference for one option over another option. Here, we focus on measuring monetary valuation, or willingness to pay (WTP), although MPLs can also be used for other purposes, such as measuring risk preferences, and our template can be adapted for such uses. A WTP MPL consists of a list of binary choices between the two options (see below), where the monetary value of at least one of these options is varied so that the value difference between them changes strictly monotonically. A given survey may contain multiple MPLs that elicit willingness to pay between different option pairs. We denote the different possible MPLs (option pairs) by MPL1, MPL2 etc. In the electricity example from South Africa, there were three possible MPLs. We defined MPL1 as a choice between one electricity vs two electricity vouchers, MPL2 as a choice between one electricity voucher vs cash, and MPL3 as a choice between two electricity vouchers vs cash.

Options In any given MPL, the subject makes repeated binary choices between a pair of options. "Option" refers to the component that remains "static" within the MPL. The options in a given MPL may be drawn from a potentially longer list of options A, B, C, ... etc. Options can be specific goods or tasks, but also the format or time of delivery of the same commodity (including money). For example, in a time preference experiment, option A may be "an immediate payment" and option B may be "a payment in two months". In a demand experiment, option A might be money, and option B the specific good for which we are trying to elicit WTP. In the South African experiment, we had option A: one electricity voucher, option B: two electricity vouchers, and option C: cash.

MPL options The possible MPL are defined by the pair of options that they compare. Within a given MPL, these options are assigned to be MPL option 1 or MPL option 2. We recommend doing this assignment in alphabetic order to avoid repeating MPLs. For example, suppose MPL2 compares option A with option C. In this case, MPL option 1 of MPL2 is option A and MPL option 2 of MPL2 is option C. From the subject's perspective, since screen side and order within the MPL are randomized, option C may appear on the left or the right on the screen, in ascending or descending order. It is therefore not necessary to define a second MPL in which MPL option 1 is option C and MPL option 2 is option A, and doing so would potentially lead to the subject seeing the same MPL twice. The order in which the possible MPL are presented to the subject can be randomized, and the number of

MPLs the subject completes can be chosen by the user. For example, each subject may see only a subset of all possible MPLs to avoid choice overload. In MPL1 in the experiment in South Africa, MPL1 option 1 was A (one voucher) and MPL1 option 2 was B (two vouchers). In MPL2, MPL option 1 was A (one voucher), and MPL option 2 was C (cash). In MPL3, MPL option 1 was B (two vouchers) and MPL option 2 was C (cash).

Element Each option has at most T values associated with it. The elements are used to define the monetary value component of each option. The value component that varies between elements within a given option should be either made explicit or already known to the subjects. We use the convention that the elements of all options are listed in weakly decreasing order of value when inputting them into the Stata script. For example, in a time preference experiment, option A (receive money today) may have a fixed value throughout the MPL, say $100. Thus, all the elements of option A are identical. In this case, option B (receive money in two months) should have different values for each element, for example $150, $145, $140,... $100. In the South Africa electricity example, the first 5 elements for all options were worth R100 and the final 5 elements decreased in value from R99 to R60.

Binary choice Each decision within an MPL is a binary choice between one element from each of the two options, denoted by subscript $t = 1, \ldots, T$ in the main text. However, one option will appear in increasing and the other in decreasing order of value. Thus, for any two options, the MPL displays binary choices between the first and last element, the second and second to last element, etc. For example, a binary choice might be between receiving $100 today vs. $130 two months from now. In this case, $100 today is the element from Option A and $130 two months from now is the element from Option B. In a demand experiment, the binary choice might be to pay a price of $4.20 and receive a widget, or pay nothing and not receive the widget.

Value difference The value difference is the difference in monetary value between MPL option 1 and MPL option 2 for every binary choice t. We use the convention that t denotes fixed combinations of MPL options. Thus, when $t = 1$, MPL option 1 has the highest possible value and option 2 has its lowest possible value, so that the value difference is decreasing. The value difference is used in analysis to measure

S.5

willingness to pay. When inputting options into the survey module, the user needs to ensure that the value difference within binary choices (starting from the last element for one option, and the first element for the other) is strictly decreasing or strictly increasing. We show examples below.

Choice order When administering the MPL, the order in which the binary choices are shown to the subject can be varied, e.g. randomly assigned. The choice order describes the order within MPL as seen by the subject. The survey module varies whether an MPL is displayed in order $t = 1, \ldots, T$ or the reverse ($t = T, \ldots, 1$) for a given subject.[2] In the SurveyCTO module, the variable that determines choice order for each subject are named *choice_order_rand*. This variable takes on the value 1 or -1. When *choice_order_rand* $= 1$, MPL option 1 takes its highest value in binary choice $t = 1$ and MPL option 2 takes its lowest value. When *choice_order_rand* $= -1$, MPL option 2 starts with the highest value and MPL option 1 with the lowest. If each subject is given more than one MPL, the variable name will also include the MPL number at the end.

Screen side order The survey module can also vary the screen side of the options within each MPL. In the SurveyCTO module, the variables that determine the screen side for each binary choice are named *screen_side_rand*. They can take on the value 1 or -1. When *screen_side_rand* $= 1$, MPL option 1 is shown on the left hand side of the screen (and MPL option 2 on the right). When *screen_side_rand* $= -1$, MPL option 1 is shown on the right hand side of the screen (and MPL option 2 on the left).

MPL order The order in which multiple MPL are shown to the subject in case the survey instrument contains multiple MPLs per subject. The survey module can vary the order in which MPLs are shown to each subject. In this document and in the SurveyCTO module, the MPL (to the subject) is denoted by roman numerals. For example, with two MPLs shown to every subject, MPL I shown might be MPL2 and MPL II shown might be MPL4. We use roman numerals to distinguish the MPL the subject is currently completing from the binary choice currently shown (see figures). In the South Africa electricity example, each subject was given two MPLs. The order in which these were asked was randomized. If a subject was

---

[2]Users can also choose not to randomize and show all MPL in the same order. A user may wish to fully randomize the order in which binary choices appear, however this requires additional coding not currently offered by our package.

assigned to receive MPL1 and MPL3, half of the sample would first complete MPL1 and then MPL3, the others would first complete MPL3, then MPL1.[3]

Realized binary choice We assume that the MPL experiment uses real stakes, that is, one binary choice is randomly selected to be implemented or paid out. This is denoted by both the MPL order and the binary choice number. The subject only learns after the MPL questions are completed which binary choice will be implemented. For example, the subject might be told that MPL II, binary choice 3 was randomly chosen for implementation. Suppose in this binary choice the subject chose one electricity voucher worth R100, then R100 would be paid out to them.

S2.2 What is in the 1_Data_Collection folder?

Below we list the contents of the 1_Data_Collection folder, along with an overview of user-supplied inputs. This serves as a quick-start section for those who have used the module in the past or know MPL well.

- `JMS_create_surveyCTO_form.do`: The main file, a Stata do file that produces a multiple price list (MPL) in SurveyCTO. The file requires some user input which is described in detail in section S2.5.

- `JMS_MPL_randomization.do`: the randomization do file. Section S2.6 describes how the randomization is implemented and how to customize this file.

- `JMS_example_mpl_input.csv`: A sample input csv file that shows how to define the options that subjects will compare (see Section S2.4). The sample file contains the options from the JMS experiment in South Africa, shown also in Figure S3. A csv file that lists all the options that will be included in the MPL module is *required*.

- `JMS_example_sampledata.csv`: A sample csv file of preloaded data for the survey module. The preloaded data generally contains identifying information of all subjects, any stratification variables that will be used in randomization, and all variables needed for other parts of the survey. Preloading data is *optional*.

- `JMS_example_surveyCTO_form.xlsx`: sample template SurveyCTO form. We provide this file for illustration but recommend that users download a current SurveyCTO template form from a SurveyCTO server (see section S2.7). A SurveyCTO template form with the the correct Survey ID is *required*.

---

[3]Note that choice, screen side, and MPL order randomization can be overridden, see below.

Although no example is included, the user can optionally provide images that show the MPL option elements (see section S2.4).

S2.3 The survey instrument as seen by the subject

This implementation of the MPL is intended for use with tablets with touchscreens that let subjects make the binary choices themselves, either independently, or with the help of a surveyor.

Each binary choice will be shown on a separate tablet screen. The module is programmed so that a (text) question can appear at the top. Below the question, each option within a binary choice can be represented by either text (see figure S2), or an image (see figure S1). Subjects make a selection by tapping the element (text or image) they want. Tapping highlights the element. The subjects needs to tap "Next" at the bottom of the screen to confirm the selection (the "swipe" option to advance in the module is disabled).

Recall that in the South Africa experiment, we defined MPL2 as a choice between one electricity voucher (here assigned to option 1) and cash (assigned to option 2). Figure S1 shows the first binary choice of the second MPL this subject is seeing (MPL II). MPL II was here randomized to be MPL2 with $choice\_order\_rand = 1$ and $screen\_side\_rand = 1$. This means that MPL option 1 has the highest value in the first binary choice and option 2 has the lowest value, so that the value difference is decreasing. In addition, MPL option 1 (one electricity voucher) is on the left hand side of the screen. The question asks, "Which would you prefer: One electricity voucher worth R100 today or R60 in cash?" Below the question, which can be read aloud by the surveyor, the pictures show a stylized electricity token with R100 value, and stylized money bills of R60 value.

All binary choices are programmed as "select_one" type fields in SurveyCTO. (See the Designing forms -> Core concepts section of the SurveyCTO Documentation for more details about field types in SurveyCTO.) The "compact" appearance option is selected, so that the options appear side by side on the screen. HTML formatting is used to increase text size and improve question layout so that it is easier for subjects to complete the MPL independently.

The exact layout in SurveyCTO may differ based on the length of the text and size of images used (if any). It is recommended that users test the survey form and adjust components such as text size so that the layout works on the devices that will be used. Figure S2 shows how binary choices are displayed on screen if no images are included.

Inputs that create the SurveyCTO form are described further in section S2.5.

Once the respondent has answered all binary choices, one of the binary choices is randomly picked to be implemented. The SurveyCTO module is programmed to show an animation of a coin flip to indicate that the tablet is making a random selection. The tablet then shows which choice was selected, and on the next screen will display the element the subject chose in this binary choice. We provide a randomization file (described in section S2.6) that places equal probability on the selection of each binary choice presented to the subject, but users can create their own randomization files to assign different probabilities.

S2.4 Defining MPL entries in the input csv file

When creating the MPL, the user needs to provide the full list of elements for all options (A, B, C,..., etc). The elements in each option along with the monetary values associated with each element are entered into a comma-separated text file. Users will later define MPL option 1 and MPL option 2 for each MPL (see above) from the list of options A, B, C,... . Each option X allows three inputs:

- The wording for each element that makes up a binary choice is entered into column *optionX*,
- the monetary value for each element that will be used in analysis is entered in column *value_optionX*, and
- optionally, the file name of the image associated with each element that may be shown on the tablet screen is in the column *image_optionX*.

If there are no images, the image column can be excluded or left blank. In this case, the options will appear on the tablet using the text provided in the column *optionX*.

For each option, the rows describe the elements of the MPL. All options need to have the same number of elements. The user must ensure that the entries are structured so that the value of at least one option changes for each binary choice. The first row should always show the highest value associated with this option. In each MPL, if $choice\_order\_rand = 1$, MPL option 1 will be shown starting at $t = 1$ and MPL option 2 will start at $t = T$, and the reverse if $choice\_order\_rand = -1$.

The template file `JMS_example_mpl_input.csv` gives an example based on the South African electricity experiment. The example has three options, reproduced in table S3. Here, we set up the MPL so it varied the value of only one of the two options, while

Figure S3: Template MPL input from the South African electricity experiment.

optionA	value_optionA	image_optionA	optionB	value_optionB	image_optionB	optionC	value_optionC	image_optionC
One electricity voucher worth R100 today	100	elect1_100.jpg	Two electricity vouchers, one worth R40 today and one worth R60 in 3 days time	100	elect2_100.jpg	R100 cash	100	cash_100.jpg
One electricity voucher worth R100 today	100	elect1_100.jpg	Two electricity vouchers, one worth R40 today and one worth R60 in 3 days time	100	elect2_100.jpg	R100 cash	100	cash_100.jpg
One electricity voucher worth R100 today	100	elect1_100.jpg	Two electricity vouchers, one worth R40 today and one worth R60 in 3 days time	100	elect2_100.jpg	R100 cash	100	cash_100.jpg
One electricity voucher worth R100 today	100	elect1_100.jpg	Two electricity vouchers, one worth R40 today and one worth R60 in 3 days time	100	elect2_100.jpg	R100 cash	100	cash_100.jpg
One electricity voucher worth R100 today	100	elect1_100.jpg	Two electricity vouchers, one worth R40 today and one worth R60 in 3 days time	100	elect2_100.jpg	R100 cash	100	cash_100.jpg
One electricity voucher worth R99 today	99	elect1_99.jpg	Two electricity vouchers, one worth R39 today and one worth R60 in 3 days time	99	elect2_99.jpg	R99 cash	99	cash_99.jpg
One electricity voucher worth R97 today	97	elect1_97.jpg	Two electricity vouchers, one worth R37 today and one worth R60 in 3 days time	97	elect2_97.jpg	R97 cash	97	cash_97.jpg
One electricity voucher worth R92 today	92	elect1_92.jpg	Two electricity vouchers, one worth R32 today and one worth R60 in 3 days time	92	elect2_92.jpg	R92 cash	92	cash_92.jpg
One electricity voucher worth R80 today	80	elect1_80.jpg	Two electricity vouchers, one worth R20 today and one worth R60 in 3 days time	80	elect2_80.jpg	R80 cash	80	cash_80.jpg
One electricity voucher worth R60 today	60	elect1_60.jpg	Two electricity vouchers, one worth R0 today and one worth R60 in 3 days time	60	elect2_60.jpg	R60 cash	60	cash_60.jpg

keeping the value of the other option constant at the maximal value of R100 (see figure S3). A user choosing this format needs to take care to enter $T/2$ invariant values and then $T/2$ strictly lower, declining values, so that the value difference varies monotonically from one binary choice to the next. We recommend this format because the maximal value available in each row is the same.

Note that some users might choose to include a binary choice in the middle where the two options have the same value, for example, an electricity voucher worth R100 and R100 in cash. At this point, the value difference is zero, whereas it is strictly positive/negative at the binary choices before/after. This may be useful when testing indifference, and in this case the MPL would have an uneven number of elements.

In our example from South Africa, the first five elements have constant value of R100, while the last five elements have strictly decreasing values below R100. In the binary choices in an MPL for this example, one option will thus start out with the highest possible value (100) while the other option will increase in value from 60 to 99 until reaching 100, at which point the other option will begin to decrease from 99 to 60.

S2.5   SurveyCTO programming in Stata

The SurveyCTO form is created using a Stata do file, `JMS_create_surveycto_form.do`. This do file draws on the input csv file described above. Moreover, users need to edit some "global" macros in this do file:

- **Name of the input csv file.** The name of the input file defining the MPL options. The default name of the file is "JMS_example_mpl_input.csv". For example: `global INPUT JMS_example_mpl_input.csv`.

- **Definition of the MPLs.** The user needs to define which options from the input spreadsheet will become MPL option 1 and MPL option 2 for each MPL that can potentially be presented to a respondent. For example, with three options A, B, and C, there are maximally three possible MPLs (pairs of options), AB, BC, and AC, but the user could specify only two, say, AB and AC. In order to specify that MPL1 should have option A as MPL option 1 and option B as MPL option 2, the global should be defined as: `global mpl1 AB`.

  Users should avoid defining options in `JMS_example_mpl_input.csv`, or combinations of options in these macros, that will never be asked. Such unused components will increase the complexity and therefore processing time for the SurveyCTO module. In the analysis, we always estimate the WTP for MPL option 1 relative to MPL option 2. The randomization file (described below) determines in what order the MPLs will appear to subjects.

- **Number of MPLs per respondent.** The user needs to define the total number of MPLs each respondent should be asked. This is an integer between 1 and $C(X, 2)$ $= \frac{X!}{(X!-2!)(2!)}$, specified in Stata as `global NO_MPL`. Which MPL from the list defined in the Stata macros above are shown to the individual respondent will be determined by the randomization file or SurveyCTO's randomization function (see below).

- **Name of the file containing randomization variables.** This is the name of an optional csv file that pre-specifies the random assignment of parameters to subjects in the MPL experiment. The default is `JMS_example_sampledata_rand.csv`. We provide a template do file to create the randomization variables to be included in this preloaded data file (more below). For example: `global RAND_DATA JMS_example_sampledata_rand.csv`.

- **SurveyCTO template form**. This is the blank survey form that will be edited to include the MPL survey questions. It is an .xlsx file that will be uploaded to the SurveyCTO server and must have the following tabs: Survey, Choices, Settings. The Stata do file will export the SurveyCTO MPL form to this file by appending the output to this file. We suggest that users download the most recent template from the SurveyCTO website. A template survey form is also copied into the GitHub repository as a sample for the user. If the GitHub SurveyCTO form is used, it should be renamed to match the form ID of the SurveyCTO form on the server that will be deployed. See the Introduction in the SurveyCTO documentation under "Core

concepts" for more information. The Stata file copies the file from the "Input" to the "Output" folder and then appends output to it. If any changes are made to the MPL and the file `JMS_create_surveyCTO_form.do` is rerun, the blank input form will be used and the output form will be overwritten. The blank form is specified in Stata as `global SCTO_FORM JMS_example_surveyCTO.xlsx`.

- **Question phrasing.** The question posed to the subject in each binary choice combines several inputs. In the electricity example, the full question reads "Which would you prefer: One electricity voucher worth R100 today or R60 in cash? Tap one of the pictures to make your choice." The wording for the options "One electricity voucher worth R100 today" and "R60 in cash" is imported from the *optionX* column from the file `JMS_example_MPL_input.csv`. The rest of the phrasing can be adjusted using the three globals Q_START, Q_OR, and Q_END.

  - Q_START contains the first part of the question. In our example, we would define: `global Q\_START ''Which do you prefer: ''`
  - Q_OR is the separator between the two options. In our example, we define: `global Q\_OR '' or ''`
  - Q_END contains the end of the sentence. In our example, the Stata do file defines:
    `global Q\_END ''Tap one of the pictures to make your choice.''`
    By default, a question mark is added after the text representing the option on the right hand side of the screen (here "R60 in cash"). The text defined in global Q_END is preceded by a line break and starts on a new line as seen in figure S1.

S2.6   Randomization

This implementation of an MPL experiment allows the user to randomize several elements that may frame the available options. An advantage of randomization is not only that any order biases are balanced on average, but also that the estimation procedure described in the main text of JMS can explicitly estimate the resulting shifts in measured WTP.

The aspects of the survey module that can be randomized, and the variables corresponding to these, are the MPL order seen by the subject, the choice order within the MPL, and the screen side within each binary choice (see glossary).

In SurveyCTO, survey questions can be shown in varying order by programming all possible realizations of the questionnaire, labeling the questions with numbers, and then assigning these numbers to subjects. The randomization in this step can be implemented in one of two ways: it can be carried out on the spot using SurveyCTO's internal randomization engine, or it can be pre-specified in an external file.

We recommend that randomization be carried out outside of SurveyCTO and pre-specified. Pre-specified randomization can improve balance or implement more complex assignment schemes. A set of random numbers for a given subject can be preloaded, and the "relevance" column in the SurveyCTO module is then used to filter which set of questions gets shown. SurveyCTO's built-in randomization cannot ensure that random assignment is balanced and that no respondent is asked the same MPL twice. Here, we describe the structure of this external randomization file. We also provide a template Stata do file that can create the randomization file for the user.

S2.6.1 Randomization variables

A data file including the randomization variables will be uploaded to the SurveyCTO server, along with the SurveyCTO form. Any preloaded data in SurveyCTO must be in csv format. The preloaded files also includes IDs and hould contain any other variables that prepopulate any fields in the survey, because SurveyCTO allows only one preloaded data set per survey. More information about preloaded data can be found in SurveyCTO documentation under "Advanced topics".

We assume that the subject sees M MPLs and each MPL contains T binary choices. The following variables describing the randomization must be included in the preloaded randomization data file (in the example this is `JMS_example_sampledata_rand.csv`):

- Order within MPL: there should be M choice-order variables indicating order within MPL. These variables can take on the value 1 or -1 and must be named `choice_order_rand_m`.
- Screen side: there should be $T \times M$ screen side variables - one for each binary choice in each MPL. These variables determine on which side of the screen option 1 and 2 appear. They can take on the value 1 or -1 and must be named `screen_side_rand_t_m`.

Taken together, there will be a total of $M + TM + M = M(2 + T)$ variables in this file.

S2.6.2 Creating the randomization file

We provide a Stata do file (`JMS_MPL_randomization.do`) to create the randomization variables. The randomization procedure requires a pre-defined sample consisting, at a minimum, of a list of pre-defined IDs. The randomization variables will be appended to each ID.

S2.6.3 Randomization procedure

The randomization do file uses permutation randomization to assign MPLs to subjects. This ensures that all the randomized components of the MPL are balanced. This can be done for the sample as a whole, or within predefined strata.

The order randomization occurs in three stages and is turned on by default, but can be de-activated at each stage. In the first stage, we assign MPLs and the order in which they are assigned to subjects. Within each stratum, the program attempts to assign each possible combination and order of MPLs to the same number of respondents. If the number of subjects in a stratum is not divisible by the number of possible assignments, the remaining subjects in that stratum will be assigned a random selection (without replacement) of MPL assignment. The variables `rand_mpl_1`, `rand_mpl_2`, etc. determine which MPL is shown to the subject first, second, and so on (what we call MPL I, MPL II, and so on above).

In the electricity example, each respondent was given 2 out of 3 possible MPLs (MPL1 (AB), MPL2 (AC), and MPL3 (BC)), and the MPLs each subject received were partly determined by a previous intervention, although the order in which they were asked was randomized. (The randomization variables would have to be manually adjusted to accommodate the externally assigned MPL).

In the second stage, the choice order is randomly assigned within stratum and MPL. Permutation randomization is used so that there is an equal number of ascending and descending orders for each MPL. Note that the user may wish to fully randomize the choice order rather than just randomizing between ascending and descending values. This requires further coding beyond what is offered by our package.

Finally, the screen side can be randomized. Within each MPL, there will be an equal number of binary choices with MPL option 1 and with MPL option 2 on the left side of the screen. Note that this does not ensure balance within binary choice across subjects. For example, by chance, it could be that in binary choice 1, a larger number of subjects see option 1 on the left side of the screen. A summary of the randomization will be printed

so that users can decide if they wish to re-randomize with a different random seed.

S2.6.4   User-provided inputs in the randomization do file

Some inputs into the randomization do file are the same as those already entered into the main survey form creation file as described in section S2.5, specifically the name of the input csv file (e.g. `global INPUT JMS_example_mpl_input.csv`), the definition of the MPLs (e.g. `global mpl1 AB`), and the number of MPLs per respondent (e.g. `global NO_MPL 1`).

In addition, the user must specify:

- **Name of the csv file containing the pre-defined sample.** This file includes the unique identifier and stratification variables as well as any other variables that will be used to pre-populate sections of the SurveyCTO form. For example, if a surveyor is revisiting a subject and there are screening questions to ensure that the correct person is being surveyed (e.g., name and address), these would also be included in the pre-defined sample file. This file must include the csv extension. Stata input: `global RAND_DATA JMS_example_sampledata.csv`

- **Random seed:** needed to ensure reproducibility.

- **Turning randomization on or off** This dofile allows users to randomize the order of the binary choices, the screen side and the order in which MPLs are displayed. Users can decide whether to turn these on or off individually.

- **Name of identifying variable in the dataset.** There must be a unique identifier for each subject. This should be a variable that uniquely identifies a subject. Stata input: `global ID id`

The following input is optional:

- **Stratification variables.** The names of the stratification variables included in the pre-defined sample data file. Users many for example want to stratify by location, income level, etc. so that there is balance between these groups. If there are no strata, these global macros should be left blank. Stata input: `global STRATA strata1 strata2`

S2.7   The Survey CTO form produced by the Stata script

The SurveyCTO form is exported as .xlsx and can be imported immediately to the SurveyCTO server. Alternatively, it can be uploaded as a Google form and then loaded to the SurveyCTO server. The user also needs to upload any preloaded data (including randomization variables, see above), as well as images representing the elements of the options, if any. The survey form can be deployed as-is with just the MPL, or users can add additional survey questions to the form.

SurveyCTO forms have three tabs that can be edited: (1) survey, (2) choices, (3) settings. This package edits the first two of these tabs in the user-provided template form (where form title and ID match the Form ID on the SurveyCTO server). The "survey" and "choices" tabs are described in detail below.

S2.7.1   Survey tab

There are three parts to the output in the "survey" tab. The first part is the "calculations" section (so named because it contains SurveyCTO fields of type "calculate"). The entries in this section determine which MPLs are displayed to each subject and in what order the randomized elements are shown. In the second part, all the binary choices for all possible MPL assignments are defined. In the last part, one of the binary choices the subject saw will be randomly selected to be implemented, and the element that was chosen by the subject in that choice is displayed to the subject.

Calculations   The section begins with a survey question that asks for a subject identifier, which allows SurveyCTO to pull the correct randomization variables from the randomization file. The default is that the subject ID is entered as text. The next part uses the uploaded randomization file to define the variables that are used in the "relevance" column of the SurveyCTO form to decide whether a given survey question will be shown to the subject or not. As an example, figure S4 shows the "calculation" section that assigns values to 10 parameters rand_side_1_1 to rand_side_1_10. These parameters determine on which side of the screen MPL option 1 and 2 will appear.

Defining binary choices   SurveyCTO can only display questions in the order in which they appear in the survey form. Therefore, MPLs need to be specified in all possible orders they can be shown to a subject. Figure S5 shows the structure of a SurveyCTO form where there are 3 possible MPLs combining options A, B, and C (MPL 1-3) and

Figure S4: SurveyCTO form calculations.

type	name	label	hint	default	appearance	constraint	constraint message	relevance	disabled	required	required message	read only	calculation
calculate	rand_side_1_1												pulldata('data.csv', 'rand_side_1_1', 'id', ${id})
calculate	rand_side_2_1												pulldata('data.csv', 'rand_side_2_1', 'id', ${id})
calculate	rand_side_3_1												pulldata('data.csv', 'rand_side_3_1', 'id', ${id})
calculate	rand_side_4_1												pulldata('data.csv', 'rand_side_4_1', 'id', ${id})
calculate	rand_side_5_1												pulldata('data.csv', 'rand_side_5_1', 'id', ${id})
calculate	rand_side_6_1												pulldata('data.csv', 'rand_side_6_1', 'id', ${id})
calculate	rand_side_7_1												pulldata('data.csv', 'rand_side_7_1', 'id', ${id})
calculate	rand_side_8_1												pulldata('data.csv', 'rand_side_8_1', 'id', ${id})
calculate	rand_side_9_1												pulldata('data.csv', 'rand_side_9_1', 'id', ${id})
calculate	rand_side_10_1												pulldata('data.csv', 'rand_side_10_1', 'id', ${id})

each subject is given 2 of the 3 MPLs (MPL I and II from the subject's perspective). For MPL I, all three MPLs are included in the survey form, even though only one of them will be seen by each subject (which one depends on the "relevance" column). Within each MPL, the binary choices can be shown so that MPL option 1 appears either in ascending or descending order (and therefore option 2 appears in descending and ascending order, respectively). They are written out in both ways, and once again the relevance column determines which of these is displayed to each subject. Finally, within each binary choice, MPL option 1 can be on the left or right hand side of the screen, and once again both of these possibilities are written out and one is selected based on the parameters defined earlier and the "relevance" column. This is repeated for MPL II.

Figure S6 show the SurveyCTO form that outputs the binary choices shown in figure S1. In this case the electricity voucher was displayed on the left hand side of the screen and only $mpl2\_order1\_side1\_choice1\_1$ will be displayed because the variable $rand\_side\_1\_1$, which was defined by the calculation section, equals 1. The two calculation rows in figure S6 are relevant if this binary choice is randomly selected to be paid out so that the wording of the chosen element is displayed to subjects and not the value assigned to that option.

S2.7.2 Choices tab

The "choices" sheet of a SurveyCTO form specifies the answer options for multiple choice questions. Here, it contains the options for each binary choice. These can be viewed as images (as seen in figure S1) or as text (figure S2). Figure S7 shows the choices tab for the binary choices in figure S6.

The choices sheet specifies the following fields:

- list_name: This field provides a name for the list of available multiple choice answers. For example, it can be seen in figure S6 that the question named "select_one

Figure S5: Structure of the MPL module in SurveyCTO.
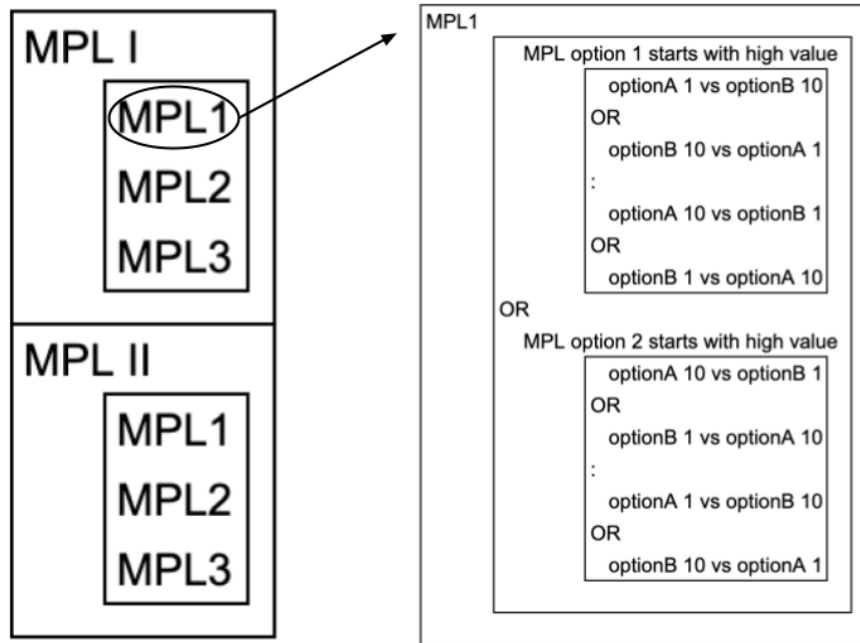
mpl2_side1_1" requires subjects to choose one of the the options in a list named "mpl2_side1_1". This list is defined in the choices tab as shown in figure S7 and provides the options A (one electricity voucher) and C (R60 in cash).

- value: The value that will be recorded as the multiple choice response in the data. The values here correspond to the options defined in the input csv (described in section S2.5), that is, option A, B, etc.

- label: The text in the label field will be displayed to the subject if there are no images. When a binary choice is randomly selected to be implemented at the end, the label will also displayed so that subjects can see what will be paid out or offered.

- image: If images are being used, the images specified here will be displayed as the MPL option.

The values, labels, and images correspond to the MPL options in the input csv file `JMS_example_MPL_input.csv`.

Figure S6: Excerpt of the SurveyCTO form "survey" tab.

type	name	label	hint	default	appearance	constraint	constrai	relevance	disabled	required
begin group	g_mpl2_choice_order1_1							${rand_dir_1} = 1		
begin group	g_mpl2_order1_choice1_1				field-list					
select_one mpl2_side1_1	mpl2_order1_side1_choice1_1	<p><h1 style = "text-align: center;"> ⓐ</h1></p><p>Which do you want to get:One electricity voucher worth R100 today -- or R60 cash? Tap one of the pictures to make your choice.</p>			compact-2			${rand_side_1_1} = 1		yes
select_one mpl2_sideN1_1	mpl2_order1_sideN1_choice1_1	<p><h1 style = "text-align: center;"> ⓐ</h1></p><p>Which do you want to get:R60 cash -- or One electricity voucher worth R100 today? Tap one of the pictures to make your choice.</p>			compact-2			${rand_side_1_1} = -1		yes
calculate	c_mpl2_order1_side1_choice1_1									
calculate	c_mpl2_order1_sideN1_choice1_1									
select_one next	next2_order1_1_1	Please tap "Next" once you have made your decision.			quickcompact					yes
end group	g_mpl2_order1_choice1_1									

Notes: A screenshot of the Survey CTO form "survey" tab. Here, the two questions shown define the same binary choice, but once with option A shown on the left ($mpl2\_order1\_side1\_choice1\_1$), and once with option C shown on the left ($mpl2\_order1\_sideN1\_choice1\_1$).

S3 Processing the Raw Data

In this section, we describe the data format required to directly use the Stata ado program file `mplwtp.ado`, which is described in detail in section S4.

If the SurveyCTO template described in section S2 above is used, we provide a Stata do file that transforms the raw survey data into a format ready for analysis, which is described in section S3.2.

S3.1 Format of the analysis-ready data

The data used by mplwtp.ado must be in "long" format where each observation is a binary choice. Figure S8 shows an example of the data. This excerpt shows all the responses to an MPL by a single respondent.

The variables required for analyzing the data using the Stata command `mplwtp` are:

- Outcome variable: A dummy variable indicating whether the respondent chose option 1 in this binary choice (`depvar` in the mplwtp syntax). In table S8, this is the variable $wtp\_chose\_option1$.
- Value difference: The difference in the value of option 1 and option 2 in this binary choice (`scalevar` in the mplwtp syntax). In table S8, this is variable $valdiff$.

S.19

Figure S7: Excerpt of the SurveyCTO "choices" tab.

list_name	value	label	image
mpl2_side1_1	A	One electricity voucher worth R100 today	elect1_100.jpg
mpl2_side1_1	C	R60 cash	cash_60.jpg
mpl2_sideN1_1	C	R60 cash	cash_60.jpg
mpl2_sideN1_1	A	One electricity voucher worth R100 today	elect1_100.jpg

Notes: A screenshot of the SurveyCTO form "choices" tab. Here, the two lists define the multiple-choice options for the same binary choice, but shown once with option A on the left, and once with option C on the left. The first question in figure S6 uses the first list ($mpl2\_side1\_1$), the second question the second one ($mpl2\_sideN1\_1$).

- Constant: Each estimate of WTP from a specific MPL must include a constant. The estimated constant term is the WTP for option 1 over option 2 in this MPL. When multiple MPL are included in the data, the WTP for each is estimated separately using dummy variables indicating the MPL (included in `indepvarlist` in the mplwtp syntax). In the example in figure S8, there are 3 MPL that each receive a dummy variable: $WTP\_A\_vs\_B$, $WTP\_A\_vs\_C$, and $WTP\_B\_vs\_C$.
- ID variables: A subject identifier (variable id in figure S8), and an MPL identifier for every unique MPL completed (variable $subject\_mpl\_id$ in figure S8).

Other variables included in the data set:

- Binary choice identifier: a variable denoting the binary choice within an MPL equivalent to the index t. In table S8, this is the variable $binary\_choice\_no$.
- Order of MPLs: if more than one MPL is shown, mpl_order indicates the order as seen by the subject.
- MPL asked: the variable mpl_type contains the internal number of each MPL shown. It contains the same information as the dummy variables $WTP\_A\_vs\_B$, $WTP\_A\_vs\_C$, and $WTP\_B\_vs\_C$ by matching each comparison to MPL numbers 1, 2, and 3.
- Order within MPL: Indicator for whether the MPL was asked in an ascending or descending order. This variable takes the values 1 or -1 if the order bias is assumed to be symmetric (see JMS for details). In table S8, this is variable $choice\_order\_rand$ (included in `indepvarlist` in the mplwtp syntax).
- Order within binary choice: Indicator for whether option 1 was on the left hand side or right hand side of the screen. This indicator takes on the values 1 or -1

Figure S8: Example of data for analysis using `mplwtp`.

id	subject_mpl_id	mpl_order	mpl_type	binary_choice_no	wtp_chose_option1	valdiff	WTP_A_vs_B	WTP_A_vs_C	WTP_B_vs_C	choice_order_rand	screen_side_rand
55	16	2	1	6	1	−1	1	0	0	−1	1
55	16	2	1	7	1	−3	1	0	0	−1	1
55	16	2	1	8	1	−8	1	0	0	−1	1
55	16	2	1	9	1	−20	1	0	0	−1	1
55	16	2	1	10	0	−40	1	0	0	−1	−1
78	17	1	3	1	0	40	0	0	1	−1	1
78	17	1	3	2	0	20	0	0	1	−1	−1
78	17	1	3	3	0	8	0	0	1	−1	1
78	17	1	3	4	0	3	0	0	1	−1	1
78	17	1	3	5	0	1	0	0	1	−1	−1

Notes: Screenshot of the analysis-ready data after processing the SurveyCTO output.

if the screen side bias is assumed to be symmetric. In table S8, this is variable *screen_side_rand* (included in `indepvarlist` in the mplwtp syntax).

- Any other control variables or group indicators (included in `indepvarlist` in the mplwtp syntax).

S3.2 do file for preparing data for analysis

The Data-processing folder in the MPL-WTP Github repo provides the Stata do file `JMS_clean_SurveyCTO_for_MPL.do` that prepares data output from SurveyCTO for analysis in the format described above.

Data collected using SurveyCTO can be downloaded as a csv file. SurveyCTO provides a Stata do file that labels and formats the data and saves as a dta file. The do file `JMS_clean_SurveyCTO_for_MPL.do` then reshapes this data from "wide" to "long" format (in the SurveyCTO output, each observation corresponds to a respondent, whereas in the analysis-ready data each observation is a binary choice) and defines all variables necessary for analysis, including variables that describe the options and the difference in value between the options for each binary choice. The example output dataset `JMS_example_processed_mpl.dta` is a dta file that is produced by running `JMS_clean_SurveyCTO_for_MPL.do` on the example input file for the South Africa electricity data. The file includes variables related to the MPL as well as other variables that are part of the SurveyCTO form. Table S9 shows an excerpt of the variables that are created by `JMS_clean_SurveyCTO_for_MPL.do` for the South Africa example.

S3.2.1 User-provided inputs in the data preparation do file

The do file `JMS_clean_SurveyCTO_for_MPL.do` requires the following inputs:

Figure S9: Variables for analysis from the JMS South Africa example.

```
              storage   display   value
variable name type      format    label     variable label

id            str2      %9s               * Subject ID
subject_mpl_id float     %15.0g             MPL by subject ID
mpl_order     byte      %9.0g              Order of MPL w/i subject
mpl_type      byte      %10.0g             MPL type
binary_choice_no
              float     %17.0g             Binary choice number w/i MPL
wtp_chose_option1
              float     %17.0g             Option 1 value is chosen (0/1)
valdiff       float     %9.0g              value of Option 1 - Option 2
WTP_A_vs_B    float     %9.0g              Dummy for MPL A vs B
WTP_A_vs_C    float     %9.0g              Dummy for MPL A vs C
WTP_B_vs_C    float     %9.0g              Dummy for MPL B vs C
choice_order_rand
              byte      %17.0g             binary choice order (asc -1 or desc 1)
screen_side_rand
              byte      %17.0g             screen side (left 1 or right -1)
```

- **The name of the input csv file.** This is the same input file defines the options that is used in as an input to the survey form creation `JMS_create_surveycto_form.do`.

  For example: `global INPUT JMS_example_mpl_input.csv`.

- **The name of the survey data dta file.** This is the output file saved after running the SurveyCTO provided dofile that labels and formats the raw data.

  For example: `global SCTO_DATA JMS_example_surveydata.dta`

- **The definition of the MPLs** This is the same input as in the survey form creation file `JMS_create_surveycto_form.do`.

  For example: `global mpl1 AB`

- **The name of the output dta file.** The processed data will be saved as a new file.

  For example: `global OUTPUT JMS_example_processed_mpl.dta`.

S4 Stata package for analysis

The third set of files we provide carries out the MPL random effects probit. The main file, `mplwtp.ado`, is a Stata user-written ado file that defines a command to estimate willingness to pay. As input, the file takes data produced from the SurveyCTO template and data processing scripts described in sections S2 and S3, or MPL data formatted to conform with the necessary inputs for estimation.

The `mplwtp` command estimates the scaled willingness to pay for option 1 over option 2 based on a latent utility model that captures both individual and choice-specific error, using data from a randomized MPL. The main paper (JMS) describes the estimating equation, assumptions and interpretation.

The program consists of the main file, `mplwtp.ado`, which defines the mplwtp program; and the file `mplwtp_xtset.ado` which defines the program for bootstrapping standard errors, and is called by `mplwtp.ado`. The user can install these files in their ado directory, or copy them into the working folder. Stata will automatically look for them in both locations.

The ado file `mplwtp` uses the pre-programmed `xtprobit` routine for panel data to estimate this model, and rescales the coefficients and standard errors so that the coefficient on the value difference between options 1 and 2 equals 1. This is equivalent to expressing all estimates in monetary units. The program uses `xtset 'panelvar'` (declaring the data set a panel data set, with the "time periods" the choices in the MPL) and clears any panel setting afterwards. This means that `xtset` needs to be reset after use of the command.

S4.1 Syntax

The `mplwtp` command uses the syntax
```
mplwtp depvar scalevar indepvarlist [if] [in] ,
panelvar(varname) [bootstrap(string) vce(string) intp(real #) sim(real #)
checkintp(string) pause(string) rawoutput(string)]
```

Varlist inputs (required)	Description
depvar	Outcome variable: dummy for choosing option 1 (varies within MPL and subject, see Section S2 for how option 1 and 2 in an MPL are defined)

scalevar	Difference in value: (option1 - option2) (fixed across subjects for MPL choices with the same choice ID t)
indepvarlist	Dummy or dummies for each population group or set of MPL for which the WTP should be estimated (at least one required, can be a constant)
	Any other covariates hypothesized to affect willingness to pay, including bias terms (typically fixed for each subject, may vary in each binary choice)

Options (required)

panelvar	Used to xtset the panel id

Other options

Standard errors:

bootstrap	Use bootstrap method to obtain the standard error of the scaled estimates
	Text input: `yes` or `no` (default)
	If no: delta method is used
vce	Choice of standard error for `xtprobit` command, see Stata documentation

Reporting

rawoutput	Whether to show the raw output of xtprobit or not
	Text input: `yes` or `no` (default)
	If no: only scaled output is shown as default

Other inputs

sim	Number of bootstrap replications (default is 100)
intp	Number of quadrature points (default set at 12 by `xtprobit`)
checkintp	Whether to use `quadchk` command to check the `intp` point input.
	Text input: `yes` or `no` (default)
pause	Only when `checkintp(yes)`
	Pause the command after checking the intp points before continuing
	Text input: `on` or `off` (default)

The panel ID will be typically an ID at the MPL-subject level. We recommend clustering standard errors at the same level, which can be achieved by specifying `vce(robust)`

or equivalently `vce(cluster panelvar)`.

Consider the data from the JMS South Africa example as seen in figures S8 and S9. In Stata, the data must be first set as panel data, with the MPL-by-subject ID as the individual variable (assuming that any error correlations occur within MPL) using `xtset subject_mpl_id`.

Then, the `mplwtp` command can be applied. For instance, the willingness to pay for electricity over cash when including order within MPL and order within choice bias terms would be estimated with the following command:

`mplwtp wtp_chose_option1 valdiff WTP_A_vs_C choice_order_rand screen_side_rand if mpl_type==2, panelvar(subject_mpl_id) nocons vce(robust)`

The condition `mpl_type==2` selects only MPLs comparing options A and C (labeled MPL 2), and within this set, the dummy `WTP_A_vs_C` is constant. The estimation results for the full data set would look as in column (4) of Table 2 in JMS.

The "Reporting" and "Other inputs" options are mainly intended for diagnostics. In a random effects probit, problems can sometimes arise when setting too few quadrature points for the integration in the random effects probit estimation (see Stata documentation for `xtprobit`). Turning on the `checkintp` option (potentially with the `pause` command) allows the user to verify that the number of quadrature points is sufficient. If the check fails, a higher number of integration points should be used. The number of quadrature points can be adjusted using the `intp` option.

If there are very few subjects with MSB, the user may also want to inspect the raw xtprobit output and compare bootstrapped and delta-method standard errors (see below). `rawoutput` displays the unscaled probit estimates.

S4.2   Output

`mplwtp` stores the following in e():

Scalars:	
`e(N_g)`	number of groups
`e(N)`	number of observations
`e(sigma_u)`	standard deviation of random effect
`e(sigma_eps)`	standard deviation of individual error
Macros:	

`e(cmd)`	mplwtp
`e(depvar)`	name of dependent variable
`e(warnings)`	`e(V)` is not calculated correctly when standard errors are based on the delta method
`e(se_method)`	method used to calculate standard errors
`e(title)`	title in estimation output
`e(properties)`	b V

Matrices:

`e(b)`	coefficient vector
`e(V)`	variance-covariance matrix of the estimators (note that this is not calculated when delta standard errors are used)
`e(ul)`	upper bound of the 95% confidence interval
`e(ll)`	lower bound of the 95% confidence interval

S4.3 Delta method standard errors

The inputs into the `mplwtp.ado` file include the dependent variable (a dummy that indicates whether option 1 was chosen), the `scalevar` variable that describes the value difference Δ_t between the two options (fixed for each binary choice t), and a dummy or set of dummies used to estimate average WTP (by subgroup if several dummies are used).

Letting β_0 be the constant or coefficient on the WTP dummy, and β_1 the estimated coefficient on Δ_t, the rescaled WTP in monetary terms is given by $\frac{\beta_0}{\beta_1}$. The standard errors on the rescaled coefficient can be obtained using the delta method, as follows:

$$\sigma^2_{WTP} = \begin{pmatrix} \frac{1}{\beta_1} & -\frac{\beta_0}{\beta_1^2} \end{pmatrix} \begin{pmatrix} V(\beta_0) & Cov(\beta_0, \beta_1) \\ Cov(\beta_0, \beta_1) & V(\beta_1) \end{pmatrix} \begin{pmatrix} \frac{1}{\beta_1} \\ -\frac{\beta_0}{\beta_1^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{V(\beta_0)}{\beta_1} - \frac{\beta_0 Cov(\beta_0,\beta_1)}{\beta_1^2} & \frac{Cov(\beta_0,\beta_1)}{\beta_1} - \frac{\beta_0 V(\beta_1)}{\beta_1^2} \end{pmatrix} \begin{pmatrix} \frac{1}{\beta_1} \\ -\frac{\beta_0}{\beta_1^2} \end{pmatrix}$$

$$= \frac{V(\beta_0)}{\beta_1^2} - \frac{\beta_0 Cov(\beta_0, \beta_1)}{\beta_1^3} - \frac{\beta_0 Cov(\beta_0, \beta_1)}{\beta_1^3} + \frac{\beta_0^2 V(\beta_1)}{\beta_1^4}$$

$$= \frac{V(\beta_0)}{\beta_1^2} - \frac{2\beta_0 Cov(\beta_0, \beta_1)}{\beta_1^3} + \frac{\beta_0^2 V(\beta_1)}{\beta_1^4}$$

In Stata and the `mplwtp.ado` file, we estimate the rescaled coefficients and associated standard errors using the `nlcom` command.

S4.4 Bootstrapped standard errors

The default method to obtain standard errors is the delta method. However, the variance of the error term is determined by the frequency of (detectable) choice specific errors (i.e., the presence of "multiple switching behavior", MSB). If MSB is rare, the variance of the error is very small, and this can lead to issues (this is another way of saying that, without observable error, a random utility model is not the right choice, and the user may consider using the `intreg` command instead if there is no MSB).

The bootstrap is used to obtain alternative point estimates and standard errors of the re-scaled coefficients (WTP estimates). Bootstrapping can be the better approach if the sample is small, or when the variance of the individual error term is small. A comparison of the bootstrapped and non-bootstrapped standard errors can reveal related issues.

For the bootstrapped standard errors, it is recommended to set the random seed before calling `mplwtp.ado`.