

## RESEARCH ARTICLE

WILEY

# Investigating (sequential) unit asking: An unsuccessful quest for scope sensitivity in willingness to donate judgments

Maximilian Maier<sup>1</sup>  | Lucius Caviola<sup>2</sup> | Stefan Schubert<sup>3</sup> | Adam J. L. Harris<sup>1</sup> 

<sup>1</sup>Department of Experimental Psychology, University College London, London, UK

<sup>2</sup>Department of Psychology, Harvard University, Cambridge, Massachusetts, USA

<sup>3</sup>Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK

## Correspondence

Maximilian Maier, University College London, Department of Experimental Psychology, 26 Bedford Way 129-B, WC1H 0AP London, UK.

Email: [m.maier@ucl.ac.uk](mailto:m.maier@ucl.ac.uk)

## Funding information

MM received a PhD scholarship by Longview Philanthropy

## Abstract

People exhibit scope insensitivity: Their expressed valuation of a problem is not proportionate with its scope or size. To address scope insensitivity in charitable giving, Hsee et al. (2013) developed the (Classical) Unit Asking technique, where people are first asked how much they are willing to donate to support a single individual, followed by how much they are willing to donate to support a group of individuals. In this paper, we explored the mechanisms, extensions, and limitations of the technique. In particular, we investigated an extension of the technique, which we call Sequential Unit Asking (SUA). SUA asks people a series of willingness-to-donate questions, in which the number of individuals to be helped increases in a stepwise manner until it reaches the total group size. Across four studies investigating donation judgments (total  $N = 2045$ ), we did not find evidence that willingness to donate (WTD) judgments to the total group increased with larger groups. Instead, our results suggest that Unit Asking (sequential or classical) increases donation amounts only through a single one-off boost. Further, we find evidence in three out of four studies that the SUA extension increases WTD judgments over Classical Unit Asking. In a fifth study ( $N = 537$ ) using a contingent valuation design (instead of donation judgments), we find scope sensitivity using all asking techniques. We conclude that, while it is difficult to create scope sensitivity in WTD judgments, SUA should be considered a promising approach to increase charitable donations.

## KEYWORDS

boosting, charitable giving, judgment and decision making, scope insensitivity, unit asking

*Everybody to count for one, nobody for more than one* (Jeremy Bentham, as quoted by Mill, 1879, p.112)

Even though many people agree with the above premise of impartiality, in practice, people usually do not help more as the size of a problem increases. This scope insensitivity is a well-studied limitation of human judgment, which describes the phenomenon that people do not scale their valuation of a quantity in proportion to its size or scope (Kahneman et al., 1999). It is commonly discussed in the

context of contingent valuation studies, which investigate how people value different goods (Baron & Greene, 1996; Desvousges et al., 1993; Kahneman & Knetsch, 1992; Lopes & Kipperberg, 2020; Veisten et al., 2004). It is additionally relevant to charitable giving (Västfjäll & Slovic, 2020), which is the primary focus of the current research. Prior work, for example, has found that when participants were asked how much money they would donate to buy Christmas presents for 20 children, they donated no more than those asked to

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Behavioral Decision Making* published by John Wiley & Sons Ltd.

help a single child (Hsee et al., 2013). Scope insensitivity has also been well-documented as a cause of neglecting human lives, explaining, for example, the relative indifference in response to some genocides (Cameron & Payne, 2011; Dickert et al., 2012, 2015; Slovic & Vstfjll, 2010). The consequences of this phenomenon can therefore hardly be overestimated. While many papers discuss the problem of scope insensitivity, a relatively minor proportion of the research output has been devoted to potential interventions to overcome it. A notable exception to this is Hsee et al.'s (2013) proposal of the Unit Asking technique.

## 1 | UNIT ASKING

Unit Asking involves first asking participants how much they would be willing to donate to help one affected individual before extending to ask how much they would be willing to pay to help many people. According to Hsee et al. (2013), the underlying mechanism through which Unit Asking reduces scope insensitivity is people's desire for consistency (Ariely et al., 2003; Luisetti et al., 2011; Thomas et al., 2018). When people are first asked about one individual and in the next step asked about a larger number, their desire for consistency drives them to donate an amount that is proportional to the amount that they donated for one individual. Unit Asking has repeatedly been shown to be highly effective in increasing donations over Direct Asking (DA) (where participants are solely asked about the total number of people and not initially asked about one individual Karlsson et al., 2020; Marcinkiewicz, 2016; Simmons, 2013). For example, participants donated more than twice as much to help 20 children when they were first asked to think about how much they would donate to help one child, than when they were only asked to think about all 20 children ( $M = \$49.42$  vs.  $M = \$18.03$ , Hsee et al., 2013).

## 2 | EXTENDING UNIT ASKING TO SEQUENTIAL UNIT ASKING

Given the applied relevance of increasing charitable donations, we aimed to improve upon the effectiveness of Unit Asking. In their original paper, Hsee et al. (2013) asked about one individual and then about a larger number of affected individuals. A natural step is, therefore, to extend this "Classical" Unit Asking (henceforth CUA) by scaling up the scope sequentially (in a stepwise manner) to larger numbers of affected individuals. For example, when eliciting donations to fund Christmas gifts for 100 children, instead of asking about 1 child and then all 100, one could first ask about 1 child, then 2, and scale up by roughly doubling (e.g., 5, 10, 20, 50) until reaching 100. We expect that this Sequential Unit Asking (SUA) will increase donations in comparison with CUA, as the repeated questions should provide additional "bite" for a mechanism related to individuals' desire for consistency to exert an influence. The first goal of this manuscript is to test whether the SUA extension increases donations over CUA.

## 3 | DOES (SEQUENTIAL) UNIT ASKING MAKE PEOPLE SCOPE SENSITIVE?

We consider scope sensitivity a continuum, with complete scope insensitivity corresponding to a complete neglect of scope and maximal scope sensitivity reflecting linear proportionality (e.g., giving 100 times more to help 100 individuals than 1 individual). In between these two extremes are different magnitudes of scope sensitivity, such as logarithmic sensitivity (e.g., giving  $\log(100)$  as much to 100 individuals as to 1 individual Fechner, 1860), or an even weaker ordinal form of scope sensitivity, which only implies giving more to larger numbers of affected individuals (e.g., giving more to 100 individuals than to 1 individual). In this manuscript, we focus on testing for this weak, ordinal form of scope sensitivity. Consequently, when we use the term "scope sensitivity", it refers to ordinal sensitivity. If an asking technique does not even show this weakest form, it cannot show any of the stronger forms.

Whereas the increase in average donations through CUA has been well-established (Hsee et al., 2013; Karlsson et al., 2020; Marcinkiewicz, 2016; Simmons, 2013), it is not yet clear to what degree this demonstrates scope sensitivity on the part of participants, even the weak ordinal form. For scope sensitivity, participants should give more as the number of affected people increases, even if the number of repeated questions remains the same (e.g., after being asked about one individual, they should donate more to help 10,000 affected individuals than to help 100). Hsee et al. (2013) recognize this in the discussion section of their paper. They cite unpublished data (Hsee, Zhang, & Lu, 2013) demonstrating that people gave more to help a total scope of 100 children than 10 children under CUA (and not DA).<sup>1</sup>

In contrast to Hsee et al. (2013), an unpublished Master's thesis (Marcinkiewicz, 2016) found CUA to only give a one-off boost to WTD judgments, independent of scope. Marcinkiewicz (2016) asked participants to donate to the charity *Global Alliance for Improved Nutrition* to help children affected by a food shortage in Mali. He used a CUA technique as described in the previous section—first, asking participants how much they would be willing to donate to help one child and then how much they would be willing to donate to help 4, 20, 200, or 2000 children (this scope varied between-subjects). While participants donated more to help the group of children than a single child, the group donation did not differ between groups of 4, 20, 200, and 2000 affected children. This finding calls into question whether Unit Asking really increases scope sensitivity in willingness to donate (WTD) judgments or rather gives a single one-off boost. Consequently, more research is required to test whether CUA actually makes participants' scope sensitive. Marcinkiewicz (2016) also showed that the higher the scope, the lower people's desire to be consistent. Because SUA presents a smaller increase in scope for each step, we predict that SUA will elicit scope sensitivity where CUA does

<sup>1</sup>We infer these simple effects from the fact that Hsee et al. (2013) report an interaction between asking condition and scope but no main effect of scope. Calculating the exact  $p$ -value of the interaction,  $F(1,310) = 4.15$ ,  $p = .042$  shows that the claim about scope sensitivity was only weakly supported in Hsee et al. (2013). Hsee, Zhang, & Lu (2013) was—to our knowledge—also never published at a later point.

not. The second goal of this paper is to test whether either CUA or SUA can result in (ordinal) scope sensitivity, such that people give more to larger numbers of affected individuals.

In the following five studies, we first compare SUA, CUA, and DA (DA—only asking a single question about the full scope) and (a) replicate the beneficial effect of CUA over DA in WTD judgments and (b) show that SUA increases WTD judgments over CUA (Studies 1a and 1b). In Study 2, we test whether CUA and SUA make participants scope sensitive by varying Asking Type and Scope. Study 2 failed to find evidence for scope sensitivity. Study 3 also observed no evidence for scope sensitivity, despite seeking to provide optimal conditions for it. Study 4 switches to contingent valuation instead of a WTD judgment-dependent variable by asking participants how much money would be *needed* to help  $x$  individuals, rather than how much they would *donate* to help  $x$  individuals. In the contingent valuation task, all asking techniques (including DA) showed scope sensitivity. This scope sensitivity was enhanced using SUA, but not CUA.

## 4 | STUDY 1A

Study 1a was a nonpreregistered pilot. In this study, we aimed to replicate the benefit of CUA in comparison with DA as in Hsee et al. (2013). In addition, we investigated whether the new SUA method could increase WTD judgments over those observed via CUA.

### 4.1 | Method

#### 4.1.1 | Participants

This study received ethical approval from the University of Oxford Central University Research Ethics Committee (reference number R56657/RE001). Participants were paid \$0.24 for this 2-min study. We uploaded the study on January 14, 2021 on Positly (<https://www.positly.com/>), selecting US participants as the target group. Positly is a front-end platform that recruits MTurk participants but adds additional quality metrics (<https://www.positly.com/participants/>). Positly blocks suspicious IPs, requires high approval rates, and requires participants to consistently pass attention checks. Initially, 406 participants signed up for the study. We excluded three participants that indicated that they were 1036, 1986, and 948 years old. After these exclusions, 403 participants remained, of which 194 were female, 207 were male, and 2 indicated another gender. The mean age was 38.27 ( $SD = 11.61$ ).<sup>2</sup>

#### 4.1.2 | Design

Participants were randomly assigned to one of three experimental conditions: DA, CUA, and SUA. The key dependent variable was the

<sup>2</sup>Because of the extreme brevity of the task, we did not add an attention check in this study. However, we added an attention check in Study 3, finding that it led to no additional exclusions.

amount (in USD) participants were willing to donate to help 100 children, entered as a free response.

#### 4.1.3 | Materials and procedure

Following Hsee et al. (2013), participants were first asked to imagine that the principal of a neighborhood kindergarten had sent them an email to ask for money that would be used to buy Christmas gifts for children:

Imagine the following: Christmas is around the corner. The principal of a neighborhood kindergarten has sent you an email asking for donations. You know her personally and trust her words. The email directs you to a website with the following questions. Please answer these questions as if you were making actual donation decisions.

They were then directed to a site that described a kindergarten with 100 children from low-income families and asked participants how much they would be willing to donate:

Thanks for visiting our website. Please read the following carefully and answer the ensuing questions. Even if you are not willing to make a donation, please still answer the questions; you may simply enter \$0. You can revise your answers, and your answers will not be recorded until you move on to the next page.

Our kindergarten currently has 100 children (like the one pictured below), they are all from low-income families. And their parents have little money to buy Christmas gifts for them. We hope you can make a donation, so we can buy Christmas gifts for them.

We used 100 children instead of the 20 used in the original study, since a setup with a larger scope seemed more appropriate to test the sequential asking technique. The experiment had three conditions.

In the control condition (as in Hsee et al., 2013), participants were asked directly how much they would be willing to donate to buy Christmas gifts for 100 children:

Please think about all **100** of these children. How much are you willing to donate to help these 100 children? Please enter the amount of money you decide and agree to donate: \_\_ \$

In the CUA condition (as in Hsee et al., 2013), participants first indicated how much they would be willing to donate to help one child and only afterwards (on a separate page) asked how much they would be willing to help 100 children. In other words, we first asked the following:

Before you decide how much to donate to help these 100 children, please first think about **one** such child and answer a hypothetical question: How much would you donate to help this one child? Please indicate the amount here: \_\_ \$”

After filling in this amount, they were asked the following question on the next page:

Now please think about all **100** of these children. How much are you willing to donate to help these 100 children? Please enter the amount of money you decide and agree to donate: \_\_ \$

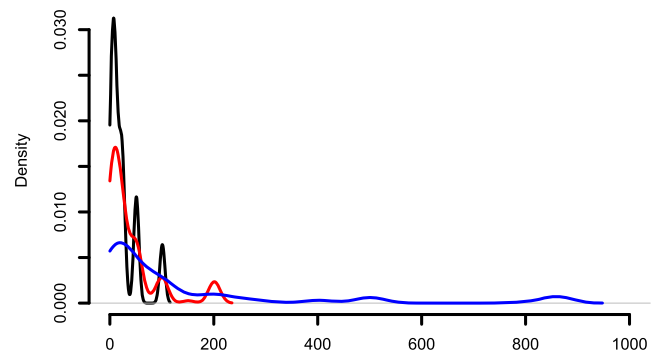
Finally, in the SUA condition (novel to this study), participants were also asked to indicate how much they would be willing to donate to help one child first (as in the CUA condition). However, instead of extending directly to 100 children, we scaled up the scope sequentially by asking (on separate pages) their WTD judgments for 2, 5, 10, 20, 50, and only then 100 children. In general, the increase per step can be determined by the  $n$ th root of the full scope, where  $n$  is the number of steps (i.e., here,  $\sqrt[5]{100}$ ). This increase is founded in Fechner's law, postulating that the subjective intensity of stimulus corresponds logarithmically to the stimulus intensity (Fechner, 1860). This law suggests that increasing the number of children by a constant multiplier will correspond to increasing by a constant sum in terms of participants' subjective stimulus intensity. We additionally round to the next multiple of five for small numbers and the next multiple of 10 for larger numbers to make the numbers more intuitive to participants (e.g., in this study, we use 50 rather than 46 as implied by the formula).

Participants provided their responses by typing any amount they saw fit (in US dollars) into a free text response box.

#### 4.1.4 | Analyses

##### *Difficulties in analyzing WTD judgments*

Open response WTD judgments often result in large outliers due to the lack of an upper bound. Previous papers on Unit Asking address this by *winsorizing* outliers and then employing a  $t$ -test or analysis of variance (ANOVA) (Hsee et al., 2013; Karlsson et al., 2020). Winsorizing replaces values above the 95th percentile of values provided with the value exactly at the 95th percentile, to reduce the skew in the data. However, winsorizing is often not effective. Even after winsorizing, the data are extremely skewed. This can make inferences overly dependent on the small subset of participants that indicate extremely large WTDs. We illustrate this using the data of Study 1a as an example. Figure 1 shows the distribution of WTD judgments in the three different groups *after* winsorizing. The distributions are still extremely skewed, and hence, the application of methods assuming normally distributed residuals is not appropriate. Winsorizing is also theoretically dubious. If a participant indicates a very high WTD judgment, it seems plausible that (1) they did not take the task seriously and should be excluded and/or (2) that they would actually



**FIGURE 1** Densities of willingness to donate (WTD) judgments in Study 1a after winsorizing. Black is the Direct Asking (DA) group, red the Classical Unit Asking (CUA) group, and blue the Sequential Unit Asking (SUA) group.

donate that much. We are not aware of any mechanism which would give rise to the idea that they meant to indicate the value that the 95th percentile participant indicated (winsorizing). We can also consider the influence of individual datapoints on the observed pattern of significance. As an example of how strongly outliers affect statistical significance patterns under winsorizing, let us consider the comparison of DA (black) and CUA (red) groups. Naively analyzing the winsorized data indicates that CUA is more effective than DA,  $t(202.95) = 3.0677$ ,  $p = .002$ . However, even though the comparison contains 271 participants and the  $p$ -value is small, the pattern of significance hinges upon the WTD judgments of six participants. If we remove the six highest values in the CUA group, the difference between the groups no longer attain the accepted significance level,  $t(226.26) = 1.916$ ,  $p = .057$ . This should give some intuition about how unstable the classical inference using winsorized data is in WTD judgments.

##### *Our analytic approach*

Due to the highlighted limitations associated with winsorizing, we take a different approach in the current manuscript, which provides Bayesian models with lognormal likelihood and nonparametric frequentist solutions to accommodate the skew in observed responses. Because we cannot meaningfully analyze scope insensitivity for participants who donated \$0 (as they would donate \$0 for any scope under scope consistency as well as scope inconsistency), and because the lognormal model cannot accommodate zeros, we excluded these participants in the following analyses. Our approach evolved during the project, and this was the preregistered approach for Study 3. This is the analysis we focus on for all studies in the manuscript. An alternative write-up with analyses exactly as preregistered is available in the OSF project. The conclusions are the same in terms of the Bayes factor categories unless stated otherwise in footnotes.

To test the differences between conditions, we specified a Bayesian analysis using the package BRMS (Bürkner, 2017, 2018). We always compared the likelihood of the data under one model assuming a difference in WTD judgments to a model assuming no difference in WTD judgments using the `bridgesampling` package (Gronau, Sarafoglou, et al., 2017; Gronau, Singmann & Wagenmakers, et al., 2017). We refer to comparisons of two groups as  $t$ -tests and to

comparisons of more than two groups as ANOVAs. Because our data were expected to be positive and quite skewed, we used a lognormal likelihood rather than the more common Gaussian likelihood ( $WTD_i \sim \text{Lognormal}(\mu_i, \sigma^2)$ , where  $i$  denotes the condition). As well as visually confirming the shape of the data, we also validated the likelihood function by comparing lognormal and normal likelihood models with leave-one-out cross-validation (Vehtari et al., 2017), which indicated superior performance of the lognormal models. In addition, the exponent of  $\mu$  should usually approximate the empirical median of the data on a linear scale, although the two can differ where the data cluster around prominent numbers (e.g., 20, 50). This is observed in our data. We present posterior medians (i.e.,  $\exp(\mu)$ ) in addition to the empirical medians in our visualization of the results.

The analysis was conditional on participants donating at all (i.e., by removing zeros), and we also removed values larger than \$10,000 (and note any instances where inferences are affected by this exclusion). We checked that our intervention did not affect the number of people donating in the first place, using a Bayesian contingency table analysis in the BayesFactor R package (Morey et al., 2015), with assumed sampling type joint multinomial and prior concentration one.

For the main analysis, we used a prior of normal ( $\mu = 3, \sigma = 1$ ) on the intercept. We used a prior of normal (0.4, 0.4) on the main effect of Asking Type. Finally, we used a prior of normal (1, 0.5) on  $\sigma$ . As the median of the lognormal distribution corresponds to  $\exp(\mu)$ , this implies that we expect a median donation of  $\exp(3) = 20.9$  in the control group and  $\exp(3 + 0.4) = 29.96$  in the intervention group. In addition, these priors result in reasonable prior predictives for this kind of donation task (see Appendix C, that is, the mean and CI predicted from the priors are similar to values we would expect).

We use Bayes factors as our primary inference criteria (Etz & Wagenmakers, 2017; Jeffreys, 1961; Kass & Raftery, 1995; Rouder & Morey, 2019; Wrinch & Jeffreys, 1921). A Bayes factor compares the probability of the data under the null (no effect) to the alternative as specified by the prior distributions outlined above. As a general rule of thumb, Bayes factors between 1 and 3 are regarded as anecdotal evidence, Bayes factors between 3 and 10 are regarded as moderate evidence, and Bayes factors larger than 10 are regarded as strong evidence (Jeffreys, 1939; Lee & Wagenmakers, 2013). The inverse of the Bayes factor can be used to describe evidence for the null hypothesis. For example, a Bayes factor between 1/3 and 1/10 would be considered moderate evidence for the null. For robustness, we also analyzed the data with frequentist nonparametric tests.

## 4.2 | Results and discussion

### 4.2.1 | Effect on proportion of participants donating and summary statistics

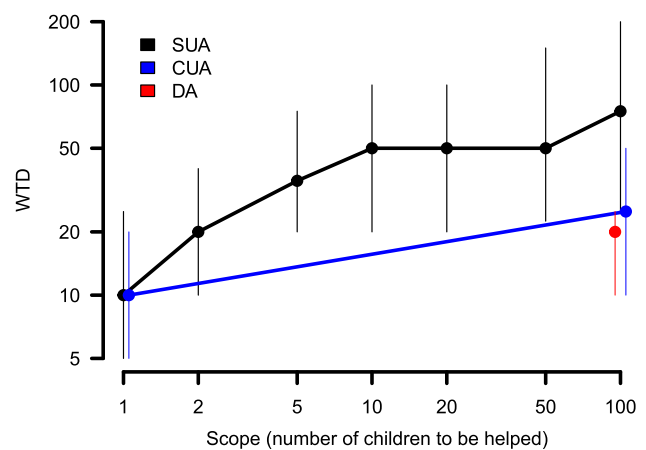
A Bayesian contingency table analysis indicated that the intervention did not affect the share of participants donating in the first place (DA: 10.37%, CUA: 13.43%, SUA: 14.18%;  $BF_{10} = 0.032$ ; see Appendix D for the share of participants donating at all across conditions in all

experiments). Our main analysis excludes WTD judgments of 0 (51 participants) and larger than 10,000 (zero participants).

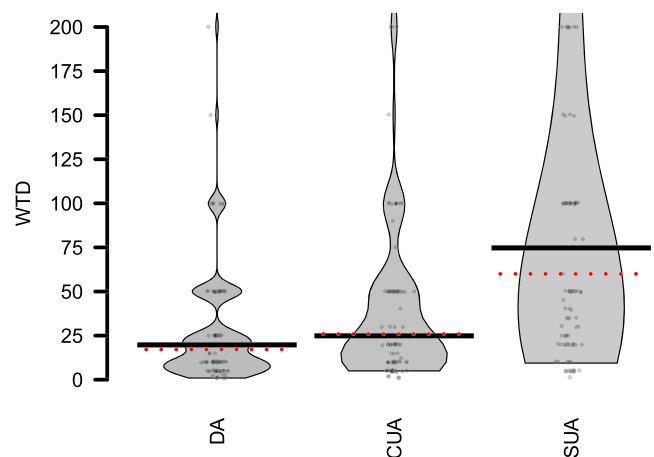
Figure 2 visualizes the median WTD judgments for all WTD questions asked (i.e., per step), whereas Figure 3 visualizes the WTD judgments only for the total scope of 100 children. The median donation for 100 children in the DA condition was \$20, the median donation for 100 children in the CUA condition was \$25, and the median donation for 100 children in the SUA condition was \$75. For one child, the median donation was \$10 in both intervention conditions. In line with the random assignment, a Wilcoxon test shows no evidence that SUA and CUA conditions differed in terms of donations for one child ( $W = 6526, p = .774$ ).

### 4.2.2 | Effect of SUA and CUA on donations

The three-level ANOVA on WTD judgments to 100 children indicated overwhelming evidence for an effect of condition ( $BF_{10} = 4.65 \times 10^9$ ).



**FIGURE 2** Median willingness to donate (WTDs) for each step in Study 1a. Error bars indicate the interquartile range.



**FIGURE 3** Willingness to donate (WTD) distribution for the full scope of 100 in Study 1a. Gray area indicates empirical density. Black lines indicate empirical medians. Red lines indicate posterior medians (i.e., the median of the posterior distribution for this parameter).

Pairwise comparisons indicated strong evidence for an effect of CUA in comparison with the DA condition ( $BF_{10} = 33.08$ ), overwhelming evidence for SUA in comparison with the DA condition ( $BF_{10} = 22.43 \times 10^{10}$ ) and strong evidence for SUA in comparison with CUA ( $BF_{10} = 22101.89$ ). The results were also corroborated with a frequentist analysis using nonparametric Wilcoxon tests (all  $p$ -values  $< .01$ ). As predicted, this study (1) replicated the effectiveness of CUA in increasing WTD judgments compared to DA and (2) showed that our extension to SUA gives a considerable additional boost over CUA.

## 5 | STUDY 1B

### 5.1 | Method

This study was a preregistered direct replication of Study 1 (<https://osf.io/uchq8>).

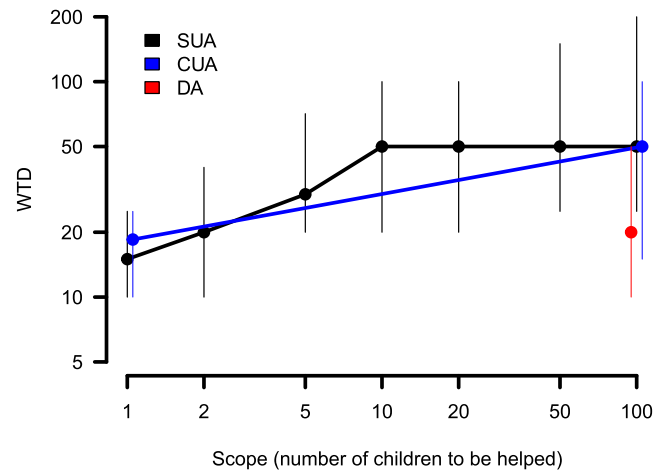
#### 5.1.1 | Participants

This study received ethical approval from Harvard University's ethics review board. We paid participants \$0.31 for this 2-min study. The study was uploaded on February 3, 2021 on Positly for US participants. Initially, 507 participants signed up for the study. We excluded one participant who indicated that their age as 3963. The mean age was 40.69 ( $sd = 12.64$ ). A total of 248 participants were female, 256 were male, 1 indicated gender "other", and 1 did not indicate their gender.

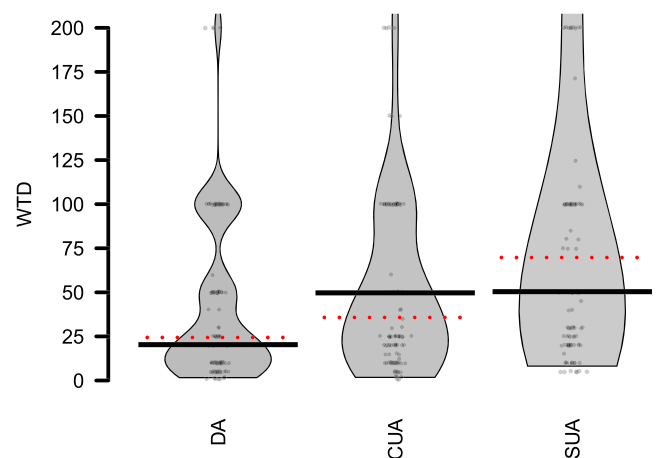
### 5.2 | Results and discussion

#### 5.2.1 | Effect on proportion of participants donating and summary statistics

A Bayesian contingency table analysis indicated that the intervention did not affect the share of participants donating in the first place (DA: 11.76%, CUA: 13.43%, SUA: 14.18%;  $BF_{10} = 0.015$ ). We excluded WTD judgments of 0 (60 participants) and larger than 10,000 (3 participants; conclusions are unaffected by this exclusion unless indicated in a footnote). We proceeded to the main analysis with the remaining 443 participants. Figure 4 displays WTD judgments per step after these exclusions. Figure 5 shows the distribution of WTDs for the full Scope of 100 children. The median donation for 100 children in the DA condition was \$20, the median donation for 100 children in the CUA condition was \$50, and the median donation for 100 children in the SUA condition was \$50. Note that if many participants indicated exactly the median, the distribution of WTD judgments between groups may still differ even though the medians are the same. Figure 5 shows that this is indeed the case, and more WTD judgments are above the median for SUA, compared with CUA, which is also reflected in the posterior medians (red line). The median donation to help one child was \$18.50 in the CUA condition and \$15 in the SUA



**FIGURE 4** Median willingness to donate (WTDs) for each step in Study 1b. Error bars indicate the interquartile range.



**FIGURE 5** Willingness to donate (WTD) distribution for the full scope in Study 1b. Gray area indicates empirical density. Black lines indicate empirical medians. Red lines indicate posterior medians (i.e., the median of the posterior distribution for this parameter).

condition. In line with the random assignment, a Wilcoxon test shows no evidence that SUA and CUA conditions differ in terms of donations for one child ( $W = 11,267$ ,  $p = .453$ ).

#### 5.2.2 | Effect of SUA and CUA on donations

The three-level ANOVA indicated overwhelming evidence for an effect ( $BF_{10} = 2.63 \times 10^{11}$ ).<sup>3</sup> Pairwise comparisons indicated moderate evidence for an effect of CUA in comparison with the DA condition ( $BF_{10} = 22.04$ ), strong evidence for SUA in comparison with the DA condition ( $BF_{10} = 18.46 \times 10^8$ ), and strong evidence for SUA in comparison with CUA ( $BF_{10} = 5\,643.68$ ). The results are corroborated

<sup>3</sup>Only 3.08 when not excluding the participants donating more than \$10,000. The change in evidence is quite large as one of the excluded participants indicated they would donate 1 million dollars. As this is likely not indicative of their actual behavior, we believe much more weight should be given to the analysis with exclusions.

with a frequentist analysis using nonparametric Wilcoxon tests (all  $p$ -values < .01).

In summary, Study 1b replicated the result that CUA increased WTD judgments over DA and that SUA additionally increased WTD judgments in comparison with CUA.

## 6 | STUDY 2

Studies 1a and 1b showed that SUA elicited higher average WTD judgments for groups of 100 than DA and CUA. However, to demonstrate scope sensitivity, we need to show that the donation is larger when more units are under consideration (e.g., people donate more to help 10,000 people than to help 100 people). Therefore, Study 2 not only varied Asking Type but also the maximum scope (100 vs. 10,000).

When scaling up SUA to a higher scope, there are fundamentally two possible approaches. First, one can keep the increase per step constant. In this case, a larger number of steps will be required to still reach the same full scope for a larger scope size (e.g., 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10,000). Second, one can keep the number of steps constant. This will result in a higher increase per step, which is required to reach a larger scope with the same number of steps (e.g., 1, 5, 20, 100, 500, 2000, 10,000). In Study 2, we investigated both of these options.

### 6.1 | Hypotheses

We preregistered three hypotheses:

1. In a  $2 \times 2$  analysis with factors Asking Type (Direct vs. CUA) and Scope (100 vs. 10,000), there will be no interaction of Asking Type and Scope (replicating Marcinkiewicz, 2016).
2. In a  $2 \times 2$  analysis with factors Asking Type (CUA vs. SUA) and Scope (100 vs. 10,000), there will be an interaction of Asking Type and Scope such that SUA minus CUA is positive and larger for Scope 10,000 than Scope 100.
3. If the effect of SUA is partially driven by an increase in the number of steps, we will observe that participants in the increase per step constant condition (SUAI) would donate more for 10,000 children than those in the number of steps constant condition (SUA).<sup>4</sup>

## 6.2 | Method

### 6.2.1 | Participants

This study received ethical approval from the Ethics Chair for the Department of Experimental Psychology, UCL (Project ID No:

<sup>4</sup>We initially preregistered to also look at an ANOVA interaction in this case; however, we realized that for Scope 100, the two Asking Types do not differ. Therefore, we used a  $t$ -test only on Scope 10,000.

EP/2021/001). We paid participants between \$0.40 and \$0.65 for participating in a 3- to 5-min study (depending on condition). The study was uploaded on August 24 and 25, 2021, targeting US participants via Positly. Based on our preregistered stopping rule, we used Bayesian sequential analysis starting with 280 participants, adding 140 additional participants in two steps until we reached 560 participants (including WTD judgments of 0). A total of 562 participants signed up, and we did not exclude any based on age. A total of 285 were female, 275 male, and 2 indicated gender “other” with mean age of 40.16 ( $sd = 12.07$ ).

### 6.2.2 | Experimental conditions

We employed a similar method to Study 1 with a number of changes. To allow more realistic scaling up to larger scopes, we replaced the kindergarten example with a vignette describing a food shortage in Ethiopia. Participants were told that in one area of the country, 100 or 10,000 children were affected by the shortage and asked how much they wanted to donate to the charity “Global Alliance for Improved Nutrition” to help the children in need (based on Marcinkiewicz, 2016). The combinations of Scope and Asking Type required to answer our Research Questions resulted in seven conditions. Participants were randomly assigned into one of the following conditions (with the exact sequence of the number of children mentioned stated in parentheses):

1. Scope 100  $\times$  DA (100)
2. Scope 10,000  $\times$  DA (10,000)
3. Scope 100  $\times$  CUA (1, 100)
4. Scope 10,000  $\times$  CUA (1, 10,000)
5. Scope 100  $\times$  SUA (1, 2, 5, 10, 20, 50, 100)
6. Scope 10,000  $\times$  SUA - number of steps constant (1, 5, 20, 100, 500, 2000, 10,000)
7. Scope 10,000  $\times$  SUAI - increase per step constant (1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10,000)

Each Unit Asking condition also included the following, exploratory, desire for consistency measure: “You said you would give  $X$  dollars for a single child and  $Y$  dollars for the group of  $N$  children. In doing so, were you trying to be consistent? That is, were you trying to allocate for each child in the group of  $N$  children as much as money as for the single child? (1 = not at all, 4 = somewhat 7 = yes, absolutely)”. The variables  $X$ ,  $Y$ , and  $N$  were adapted based on the participant's responses. The consistency measure was asked at the end of the study to avoid any influences on the key variable of interest (WTD).

### 6.2.3 | Analyses

We used the same analysis as Study 1b with the addition of a prior of normal (0.25, 0.25) on the interaction. The reason why the priors are increasingly more narrow is that, given the lognormal likelihood, the

prediction on the linear scale corresponds to the exponent of the marginal mean. Therefore, using similar priors on main effects and interactions would result in the prediction of an extremely large interaction effect on the linear scale. We tested for the presence of the hypothesized interactions by comparing models that include the interaction to models that do not include the interaction using Bayes factors. In addition, we conducted a frequentist analysis using rank-based (non-parametric) tests for main effects and interactions with the package `rfit` (Kloke et al., 2012) and independent sample Wilcoxon tests when comparing pairwise differences.

## 6.3 | Results

As we did not achieve sufficient evidence on all three critical tests in the preregistered intermittent analyses, we collected the full sample of 562, consistent with our preregistration.

### 6.3.1 | Effect on proportion of participants donating and summary statistics

A Bayesian contingency table analysis again indicated that the intervention did not affect the share of participants donating at all ( $BF_{10} = 0.19$ ; see Table D1 for proportions). Therefore, we proceeded to the main analysis excluding WTD judgments of 0 (133 participants) and larger than 10,000 (exclusions by condition: DA = 2; CUA = 2; SUA - number of steps constant = 11; SUA - increase per step constant = 4; conclusions are unaffected by this exclusion unless indicated in a footnote), leaving us with a total of 410 participants.

Figure 6 shows the median WTD judgments per step, and Figure 7 shows the distribution of WTD judgments for the total number of children (see also Table D1). In line with the random assignment, a Wilcoxon test shows no evidence that SUA and CUA conditions differ in terms of donations for one child ( $W = 10,992, p = .767$ ).<sup>5</sup>

### 6.3.2 | H1: Comparing CUA and DA

The first hypothesis, that CUA would not make participants scope sensitive, received only weak support. A Bayesian interaction test in a  $2 \times 2$  ANOVA with factors Scope (100 vs. 10,000) and Asking Type (DA vs. CUA) implied weak evidence against an interaction of Asking Type and Scope ( $BF_{10} = 0.50$ ).<sup>6</sup> In addition, there was strong evidence for a main effect of CUA versus DA ( $BF_{10} = 10.03$ ).<sup>7</sup> A nonparametric ANOVA also found no evidence for an interaction of Asking Type and

Scope,  $F(1,228) = 0.49, p = .488$ , and a significant main effect of CUA versus DA,  $F(1,228) = 5.11, p = .025$ .

### 6.3.3 | H2: Comparing SUA and CUA

We found weak evidence against the second prediction that SUA would result in more scope consistency than CUA, as revealed by an interaction between SUA (number of steps constant) versus CUA and Scope ( $BF_{10} = 0.64$ ).<sup>8</sup> As in the previous study, we found a main effect of SUA versus CUA ( $BF_{10} = 11.70$ ),<sup>9</sup> although the evidence here is much weaker. These findings were corroborated with a nonparametric, frequentist analysis finding no interaction,  $F(1,247) = 0.35, p = .552$ , and a (just) significant main effect of SUA versus CUA;  $F(1,247) = 3.96, p = .047$ . Further, when only comparing the increase in WTD judgments under SUA versus CUA for Scope 100, which is similar to Study 1, the evidence for the effect is only moderate ( $BF_{10} = 4.085$ ).<sup>10</sup> This is surprising given the strong evidence found in Study 1.

### 6.3.4 | H3: Comparing different types of SUA

We found no evidence that donations were higher for 10,000 children in the SUA1 (increase per step constant) condition than the SUA (constant number of steps) condition, as revealed by a t-test ( $BF_{10} = 0.67$ ).<sup>11</sup> This is also supported by a nonparametric Wilcoxon test,  $W = 1473, p = .949$ .

### 6.3.5 | Exploratory analyses: Investigating judgments in intermediate steps and desire for consistency

The donation patterns so far give rise to some interesting avenues for more detailed examination. Two conditions used identical sequences of steps up to 100 participant: SUA with constant increase per step for Scope 10,000, dotted gray line in Figure 6 and SUA for Scope 100, solid black line in Figure 6. This raises the question of why the WTD judgments are not higher for the full scope in SUA1 with constant increase per step for Scope 10,000, given the larger number of affected individuals as well as larger number of steps. Two explanations come to mind: (1) participants reach some kind of donation ceiling after which they do not donate more and (2) participants (who were informed in the beginning what the maximum scope would be) think ahead and donating less for one child when they know the full scope will be higher. In other words, they might first form a judgment

<sup>5</sup>The reason why the medians in the figure for one child are different even though there is no significant difference is that participants cluster their responses around prominent numbers (i.e., 10, 20, 50). Therefore, only a few participants changing their judgment can result in a jump between two prominent numbers.

<sup>6</sup>We found strong evidence under the preregistered model.

<sup>7</sup>We found only moderate evidence when not excluding WTDs larger than \$10,000 ( $BF_{10} = 6.48$ ).

<sup>8</sup>We find moderate evidence for this interaction when not excluding participants donating more than \$10,000 ( $BF_{10} = 3.61$ ) and strong evidence against this when using the preregistered model.

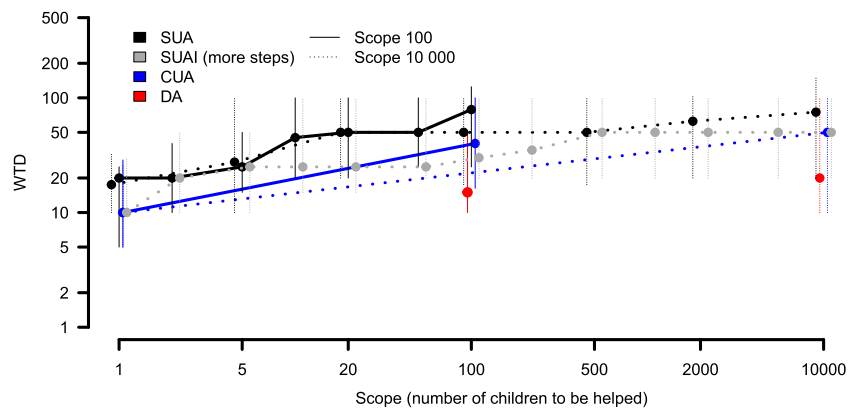
<sup>9</sup>Moderate evidence in the preregistered analysis.

<sup>10</sup>Only 2.37 without the exclusions.

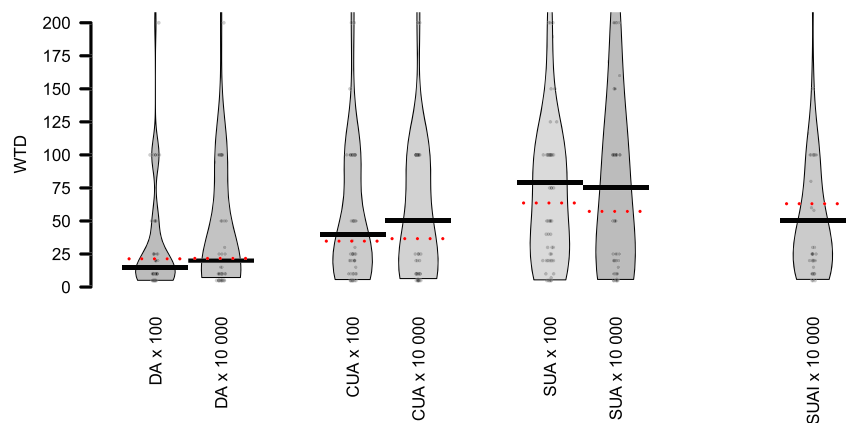
<sup>11</sup>Moderate evidence that they do not differ under the preregistered analysis.



**FIGURE 6** Median willingness to donate (WTDs) for each step in Study 2. Error bars indicate interquartile range.



**FIGURE 7** Willingness to donate (WTD) distribution for the full scope in Study 2. Gray area indicates empirical density. Black lines indicate empirical medians. Red lines indicate posterior medians. SUA I is for Sequential Unit Asking (SUA) with increase per step constant.



**TABLE 1** Share of participants showing a strong monotonic increase in scope 10,000 SUA I (increase per step constant) condition and scope 100 SUA condition.

Step	SUA - maximum scope 100	SUA I - maximum scope 10,000
1 → 2	52%	43%
2 → 5	53%	42%
5 → 10	45%	34%
10 → 20	33%	24%
20 → 50	36%	19%
50 → 100	31%	18%
100 → 200		19%
200 → 500		20%
500 → 1000		19%
1000 → 2000		11%
2000 → 5000		15%
5000 → 10,000		15%

Abbreviation: SUA, Sequential Unit Asking.

about what to donate according to the full scope and then donate a proportional fraction of this to one child. Table 1 shows the share of participants that indicate a strong monotonic increase at each step (i.e., donate more for more children). Table 1 confirms what is suggested by Figure 6. In the Scope 100 condition, more participants increase on each step in comparison with the Scope 10,000 condition. In addition, the proportion of participants increasing is lower for the

larger scopes in the Scope 10,000 condition. In other words, both looking ahead and reaching a donation ceiling may play a role in this donation behavior.

Which kinds of participants are more likely to keep increasing their donations, and which are more likely to drop out? One hypothesis is that participants that donate more for one child are more likely to drop out later, as they are more likely to run out of money. To test this, we conducted a linear regression predicting the number of strong monotonic increases from the WTD judgments for a single child. However, we did not find any evidence for the notion that WTD judgments for one child were related to monotonic increases for any of the three SUA conditions (maximum scope 100:  $F(1,79) = 2.18$ ,  $p = .501$ , maximum Scope 10,000 and increase per step constant:  $F(1,77) = .764$ ,  $p = .385$ , and maximum Scope 10,000 and number of steps constant:  $F(1,77) = .666$ ,  $p = .417$ ). Finally, we found evidence against an effect of Asking Type on our exploratory measure, “desire for consistency” ( $BF_{10} = 0.055$ ), which we included in line with Marcinkiewicz (2016). However, this might not rule out “desire for consistency” as an explanation for the effect given the likely insensitivity of this one item measure.

## 6.4 | Discussion

We again replicated the benefit of SUA over CUA; however, the evidence was much weaker than in Study 1 (though still moderate in strength).

We found no evidence that unit asking increased scope sensitivity in this study. This finding holds for SUA and CUA—both techniques only give a constant boost independent of scope. In other words, repeated asking leads people to indicate higher WTD judgments, but this effect appears *independent* of Scope (the number of children affected). Perhaps the most surprising result is that we do not observe scope sensitivity in the SUAI—increase per step constant—condition. To explain this result, we suggest that some participants might reach a donation ceiling (as suggested by the smaller number of participants displaying monotonic donation patterns for the larger number of affected individuals; see Table 1). Alternatively, some participants might seek to ensure consistency by donating less to a single child when they had an idea of what the maximum scope would be. Finally, some participants might simply find the large number of repetitive questions that were asked in this condition unpleasant or irritating and, therefore, disengage from the task. Study 3 included alternative optimal conditions for observing scope sensitivity on the basis of these conjectures.

## 7 | STUDY 3

Study 3 aimed to provide the most favorable conditions for scope sensitivity. To reduce the likelihood of participants reaching a donation ceiling, we reduced the maximum scope in this study from 10,000 to 50. Achieving scope consistency for such a smaller scope is likely more realistic, and Hsee et al. (2013, p.1806) would appear to agree: “We should also note, however, that the ability of unit asking to increase scope sensitivity is likely limited; if the target numbers are large—for example, 1000 versus 10,000—respondents may encode either number as ‘a lot’ and not differentiate the two”. To avoid participants planning forward when knowing the maximum scope, we included conditions where we do not tell participants the maximum scope in advance of the iterative procedure. Finally, we also reverted back to the kindergarten vignette that we had used in Study 1 which showed the strongest benefit of SUA over CUA. Overall, we aimed to test the following four preregistered research questions in this study:

1. Can we identify scope sensitivity for any of the asking techniques?
2. Does SUA increase WTD judgments over CUA?
3. Does SUA make participants more scope sensitive than CUA?
4. Does telling people the maximum scope beforehand affect scope sensitivity?

### 7.1 | Method

#### 7.1.1 | Participants

This study received ethical approval from the Ethics Chair for the Department of Experimental Psychology, UCL (Project ID No: EP/2021/001) and was preregistered at <https://osf.io/ezrs9>. We paid participants \$0.37 for participating in a 2-min study. The study was

uploaded between the fourth and sixth of November 2021, targeting US participants via Positly. Initially, 574 participants signed up; no participants were excluded by age, and none failed the attention check. The mean age was 40.31 ( $sd = 12.62$ ). A total of 289 participants were female, 279 were male, and 6 reported “other”.

#### 7.1.2 | Experimental conditions

To test the research questions, we used several combinations of Asking Type and Scope. Unless otherwise indicated, and unlike in Studies 1 and 2, participants were *not* informed of the maximum scope in advance. Participants were randomly assigned into one of the following conditions (with the exact sequence of steps stated in parentheses):

1. Scope 10 × CUA (1, 10)
2. Scope 50 × CUA (1, 50)
3. Scope 10 × SUA (1, 2, 5, 10)
4. Scope 50 × SUA—number of steps constant (1, 4, 15, 50)
5. Scope 10 × SUA (1, 2, 5, 10) [participants know maximum scope when answering first question]
6. Scope 50 × SUA—number of steps constant (1, 4, 15, 50) [participants know maximum scope when answering first question]
7. Scope 50 × SUAI—increase per step constant (1, 2, 5, 10, 20, 50)

We did not include a DA condition this time as, at this point, it is well-established that people are scope insensitive under DA and that both CUA and SUA increase WTD judgments relative to this baseline.

## 7.2 | Results

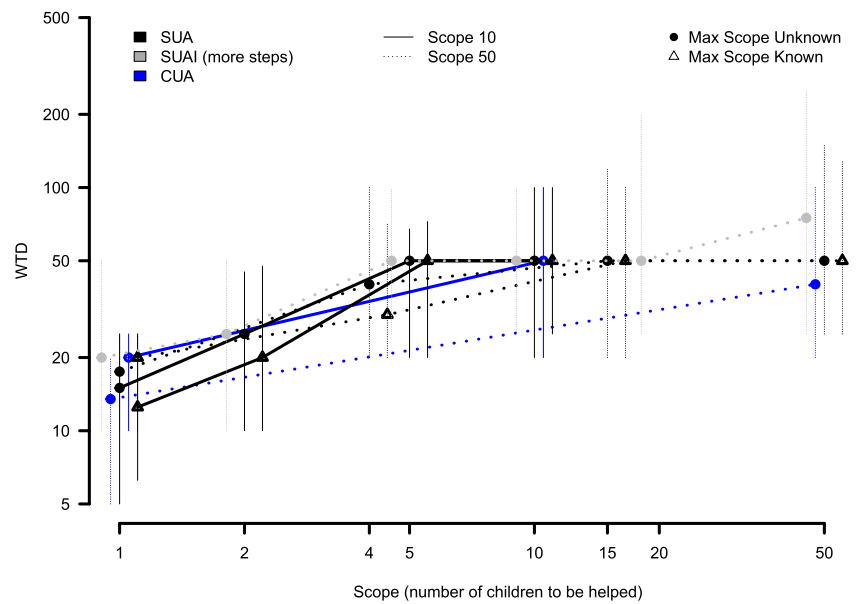
### 7.2.1 | Effect on proportion of participants donating and summary statistics

We used the same statistical model as in Study 2. Asking Type did not affect the number of participants donating in the first instance ( $BF_{10} = 0.03$ ; see Table D2 for proportions). We proceeded to our main analysis, excluding participants that donated 0 (70 participants) and participants that donated more than 10,000 (0 participants). Figure 8 visualizes the median donation trajectories for the different steps, and Figure 9 shows the distributions of WTD judgments for the full scope (see also Table D2). In line with the random assignment, a Wilcoxon test shows no evidence that SUA and CUA conditions differed in terms of donations for one child ( $W = 26,393, p = .916$ ).

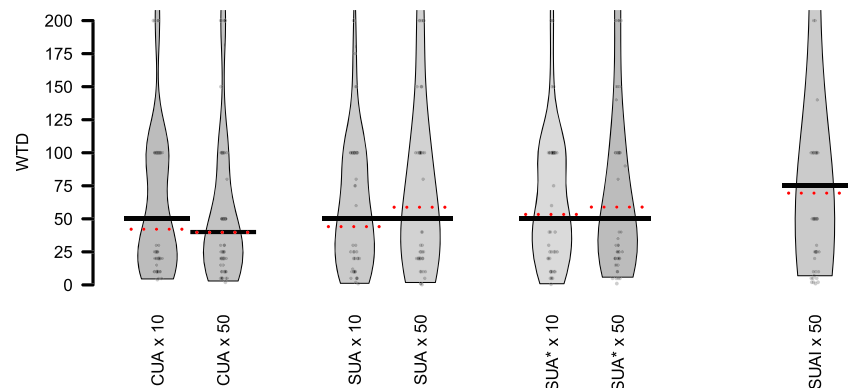
### 7.2.2 | RQ1: Can we identify scope sensitivity in any format?

We tested this with four comparisons between the WTD judgments for 50 versus 10 participants for both CUA and SUA

**FIGURE 8** Median willingness to donate (WTDs) for each step in Study 3. Error bars represent the interquartile range.



**FIGURE 9** Willingness to donate (WTD) distribution for the full scope in Study 3. \*SUA conditions where the maximum Scope was unknown to participants. In all other conditions, the maximum Scope is known. Gray area indicates empirical density. Black lines indicate empirical medians. Red lines indicate posterior medians (i.e., the median of the posterior distribution for this parameter). SUAI is for Sequential Unit Asking (SUA) with increase per step constant.



(i.e., in terms of the Experimental Conditions 1 vs. 2, 3 vs. 4, 3 vs. 7, and 5 vs. 6). We found moderate evidence that participants donate the same for 50 vs. 10 individuals for CUA ( $BF_{10} = 0.29$ ). For SUA, we compared the Scope 10 condition to the Scope 50 condition (number of steps constant) and the Scope 50 condition (increase per step constant). We found no evidence for either the null or alternative hypothesis for the number of steps constant comparison ( $BF_{10} = 0.86$ ). For the increase per step constant comparison, we found weak evidence that participants donated more with larger scopes ( $BF_{10} = 2.55$ ). In addition, for SUA (number of steps constant), we also have one condition where participants knew the maximum scope when starting the experiment. Here we find weak evidence for participants donating less on the larger scopes ( $BF_{10} = 0.36$ ). In sum, Study 3 did not reveal convincing evidence for scope sensitivity in any format. This lack of evidence is also corroborated by nonparametric Wilcoxon tests. This indicates no evidence for scope consistency under CUA ( $p = .357$ ), no evidence for scope consistency under SUA (number of steps constant,  $p = .241$ ), no evidence for scope consistency under SUAI (increase per step constant,  $p = .077$ ), and no evidence for scope consistency when telling people the maximum scope beforehand ( $p = .883$ ).

### 7.2.3 | RQ2: Does SUA increase WTD judgments over CUA?

Comparing SUA (number of steps constant) to CUA (1 and 2 vs. 3 and 4), we find no evidence for or against an effect of SUA ( $BF_{10} = 0.99$ ).<sup>12</sup> The Wilcoxon test is (just) significant in favor of SUA ( $W = 9159, p = .047$ ). When instead comparing SUAI (increase per step constant, 1 and 2 vs. 3 and 7) to CUA, we find weak evidence for a difference: Bayesian analysis:  $BF_{10} = 2.68$ , Frequentist analysis:  $W = 9919.5, p = .018$ . Note, however, that we preregistered to use SUA (number of steps constant) to test RQ2.

### 7.2.4 | RQ3: Does SUA make participants more scope sensitive than CUA?

For the interaction between Asking Type (SUA [number of steps constant] vs. CUA) and Scope (10 vs. 50; i.e., 1 and 2 vs. 3 and 4), we find no evidence that SUA increases scope consistency in comparison with CUA ( $BF_{10} = 1.17$ ). When instead testing this interaction using SUAI

<sup>12</sup>Using only SUA (number of steps constant).

(increase per step constant; i.e., 1 and 2 vs. 3 and 7), we find weak evidence that SUAI increases scope consistency in comparison with CUA ( $BF_{10} = 2.48$ ). This finding is partially in line with a nonparametric interaction test, which shows no evidence for an interaction in the number of steps constant condition,  $F(1,287) = 3.06$ ,  $p = .081$ ; however, it does show evidence for a (just) significant interaction for SUAI (increase per step constant),  $F(1,286) = 5.21$ ,  $p = .023$ .

### 7.2.5 | RQ4: Does telling people the maximum scope beforehand affect scope sensitivity?

When testing an interaction between telling people the maximum scope before and Scope (i.e., 3 and 4 vs. 5 and 6), we find no evidence that telling people the maximum scope beforehand increases scope sensitivity ( $BF_{10} = 0.42$ ) for the SUA conditions. This is also confirmed by a robust frequentist ANOVA,  $F(1,281) = 0$ .

## 7.3 | Discussion Study 3

### 7.3.1 | Effect of SUA versus CUA

#### *No evidence for a difference in Study 3*

In this study, we did not find evidence for the main effect of SUA. Importantly, we also did not find evidence *against* an effect of SUA. To investigate whether this result is driven by the smaller scope in this study, we compared the SUA versus CUA effect in this study (for the condition with 50 participants) to that in Study 1a (Scope = 100), where we found the strongest effect of SUA. We tested the interaction between Condition (SUA vs. CUA) and Study to investigate whether the difference in conditions was affected by the study. As we did not have a Unit Asking condition where the scope was known in advance in Study 3, we used the conditions where the scope is not known for this comparison. We only found weak (and nonsignificant) evidence for this interaction (Bayesian analysis:  $BF_{10} = 2.32$ ; nonparametric ANOVA:  $F(1,373) = 1.400$ ,  $p = 0.237$ ). We conclude, therefore, that the effect of SUA versus CUA does not reliably differ across the studies.

#### *Effect of SUA vs. CUA across all donation studies*

To further test the overall effect of SUA, we pooled the data from all four studies and one study presented in Appendix A, which employed a different design that likely diminished the effectiveness of SUA. We tested for an overall effect of SUA versus CUA across these five studies using a Bayesian mixed-effects model with random effects for Study and Scope. We used the same priors as in Study 3 and  $BRMs$ ' default priors on the random effects.<sup>13</sup> When pooling across studies, we only find moderate evidence for an effect of SUA in comparison with CUA ( $BF_{10} = 3.65$ ). If we only look at the studies in the main text of this manuscript, the evidence is somewhat stronger ( $BF_{10} = 4.50$ ).

<sup>13</sup> $t(3,0,2.5)$  on random intercepts and slopes and LKJ (1) on the correlation between random effects

In conclusion, when pooling across all relevant studies, there is moderate evidence that SUA results in higher WTD judgments for helping multiple individuals than does CUA.

### 7.3.2 | No evidence for scope sensitivity under CUA and SUA

We again did not observe any evidence for scope sensitivity, even after creating these most favorable conditions to observe scope sensitivity in donation judgments. We believe that the two most likely explanations for this result are as follows:

1. Even with reduced scope, participants' budget constraints limit their donation judgments to an extent that scope sensitivity may not be observed with WTD judgments.
2. None of the asking techniques promotes scope sensitivity in general.

If the first explanation is correct, we would expect to see evidence for scope sensitivity in a contingent valuation version of the study. In other words, instead of asking participants how much money they would donate to buy Christmas gifts for the children, one would ask how much money they think is required to buy Christmas presents for these children. As no willingness for a personal donation is asked for, budget constraints would no longer explain the lack of scope sensitivity. We tested such a contingent valuation setup for Study 4.

## 8 | STUDY 4

To switch to a contingent valuation setup in Study 4, we replaced the question about donations in Study 3 with the following question:

Please think about all 50 [10 in the lower scope condition] of these children. How much money do you think is needed to buy Christmas gifts for these 50 children? Please indicate the amount here: \_\_\$

We modified the background scenario from Study 3 by telling participants that they are active in the local community and occasionally give advice to people in their community when they have admin. related questions. In a next step, we informed them that the principal of a neighborhood kindergarten has contacted them to ask how much money they think is needed to buy Christmas gifts for the children. We elicited the contingent valuation judgments with the different asking techniques specified in Study 3.

The aim of this study was to test the following preregistered research questions:

1. Does CUA increase judgments over DA?
2. Does SUA increase judgments over CUA?
3. Which asking techniques induce scope sensitivity?

4. Does CUA make participants more scope sensitive than DA?
5. Does SUA make participants more scope sensitive than CUA?

## 8.1 | Method

### 8.1.1 | Participants

This study was approved by the Ethics Chair for the Department of Experimental Psychology, UCL (Project ID No: EP/2022/001) and pre-registered online (<https://osf.io/abcd12>). Participants were paid \$0.46 for their participation in a 2.5-min study. The study was uploaded on February 10, 2023 via Positly. Initially, 567 participants signed up, 1 participant was excluded because they did not indicate their age, and 28 failed the attention check (a question asking how many children were in the kindergarten). The mean age of participants was 41.26 ( $sd = 11.77$ ) with 268 being female, 265 male, and 5 reporting "other".

### 8.1.2 | Experimental conditions

To test the preregistered research questions, we randomly assigned participants to one of the following combinations of Asking Type and Scope:

1. Scope 10 × DA (10)
2. Scope 50 × DA (50)
3. Scope 10 × CUA (1, 10)
4. Scope 50 × CUA (1,50)
5. Scope 10 × SUA (1, 2, 5, 10)
6. Scope 50 × SUA—number of steps constant (1, 4, 15, 50)
7. Scope 50 × SUA—increase per step constant (1, 2, 5, 10, 20, 50)

## 8.2 | Results and discussion

### 8.2.1 | Effect on proportion of participants donating and summary statistics

The statistical model was the same as in Study 3. However, this time, none of the participants indicated an amount of zero; thus, we did not conduct the contingency table analysis and proceeded directly to our main analysis, excluding participants who indicated more than 10,000 (12 participants, 4 for each asking type; conclusions are unaffected by this exclusion unless indicated in a footnote). Figure 10 visualizes the median contingent valuation trajectories for the different steps, and Figure 11 shows the distribution of contingent valuation judgments for the full scope. The figures suggest that this time participants did increase their judgments for larger numbers of affected children. In line with the random assignment, a Wilcoxon test shows no evidence that SUA and CUA conditions differed in terms of valuations for one child ( $W = 15,195$ ,  $p = .140$ ).

### 8.2.2 | RQ1 and RQ2: Which asking technique induces the highest valuation judgments?

When comparing DA and CUA (Conditions 1 and 2 vs. 3 and 4), we find no evidence for a difference in judgments (Bayesian analysis:  $BF_{10} = 0.21$ , frequentist analysis:  $W = 11,060$ ,  $p = .653$ ).

When comparing SUA to CUA, we also find no evidence for a difference in judgments when using the number of steps constant condition (Conditions 3 and 4 vs. 5 and 6; Bayesian analysis:  $BF_{10} = 0.87$ , frequentist analysis:  $W = 10,378$ ,  $p = 0.329$ ); however, we find moderate evidence with the increase per step constant, SUA1, condition (Conditions 3 and 4 vs. 5 and 7, Bayesian analysis:  $BF_{10} = 6.13$ ,<sup>14</sup> frequentist analysis:  $W = 1093$ ,  $p = .102$ ).

### 8.2.3 | RQ3: Which asking techniques induce scope sensitivity?

When comparing donations for Scope 10 and Scope 50, we find strong evidence for scope sensitivity for all of the asking techniques:

- **DA** (Conditions 1 vs. 2): Bayesian analysis:  $BF_{10} = 3.46 \times 10^{12}$ ; frequentist analysis:  $W = 843$ ,  $p < .001$ .
- **CUA** (Conditions 3 vs. 4): Bayesian analysis:  $BF_{10} = 2.48 \times 10^6$ ; frequentist analysis:  $W = 1186.5$ ,  $p < .001$ .
- **SUA** (Conditions 5 vs. 6): Bayesian analysis:  $BF_{10} = 1.59 \times 10^{15}$ ; frequentist analysis:  $W = 771.5$ ,  $p < .001$ .
- **SUA** (increase per step constant; Conditions 5 vs. 7): Bayesian analysis:  $BF_{10} = 1.82 \times 10^{16}$ ; frequentist analysis:  $W = 724$ ,  $p < .001$ .

### 8.2.4 | RQ4: Does CUA make participants more scope sensitive than DA?

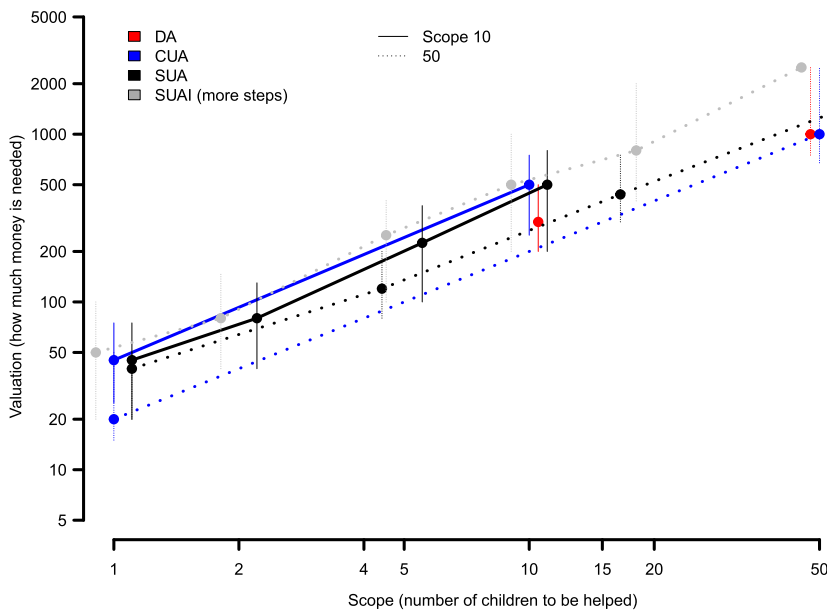
We find evidence against the hypothesis that CUA makes participants more scope sensitive than DA (interaction of Conditions 1 and 2 vs. 3 and 4; Bayesian analysis:  $BF_{10} = 0.32$ , frequentist analysis:  $F(1,298) = 0.33$ ,  $p = 0.564$ ).

### 8.2.5 | RQ5: Does SUA make participants more scope sensitive than CUA?

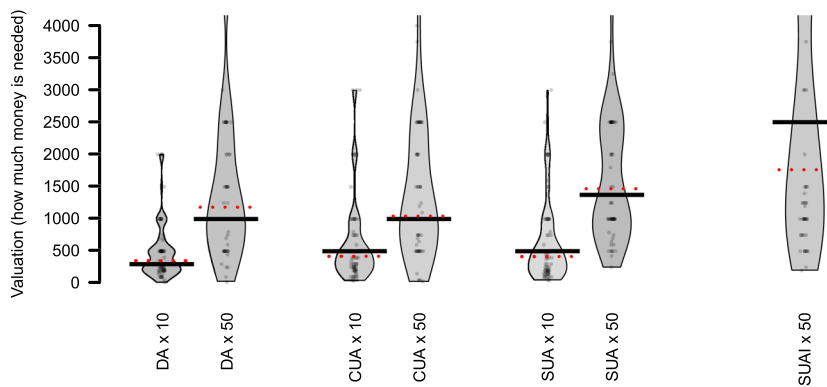
For the number of steps constant condition, we find moderate evidence that SUA makes participants more scope sensitive than CUA (interaction of Conditions 3 and 4 vs. 5 and 6; Bayesian analysis  $BF_{10} = 4.16$ , frequentist analysis,  $F(1,294) = 4.31$ ,  $p = .039$ ),<sup>15</sup> while for the increase per step constant condition, we find strong evidence for this effect (interaction of Conditions 3 and 4 vs. 5 and 7; Bayesian analysis:  $BF_{10} = 36.05$ , frequentist analysis:  $F(1,297) = 12.915$ ,  $p < .001$ ).

<sup>14</sup>Only weak evidence when not excluding donations larger than 10 000,  $BF_{10} = 2.54$

<sup>15</sup>Without the exclusion of donations larger 10,000  $BF_{10} = 2.93$ , frequentist analysis,  $F(1,294) = 2.63$ ,  $p = .106$



**FIGURE 10** Median contingency valuation judgments for each step in Study 4. Error bars represent the interquartile range.



**FIGURE 11** Contingency valuation distribution for the full scope in Study 4. Gray area indicates empirical density. Black lines indicate empirical medians. Red lines indicate posterior medians (i.e., the median of the posterior distribution for this parameter). SUAI is for SUAI (increase per step constant).

### 8.2.6 | Conclusion Study 4

For contingent valuation judgments, we only find evidence for a benefit of SUA over CUA in average judgments when using the increase per step constant condition, SUAI, rather than the number of steps constant condition. We further find evidence for an interaction between Scope and CUA versus SUA, for both SUA types. In other words, SUA may only increase contingent valuation judgments if (1) the scope is larger than 10 and (2) a sufficient number of steps is asked (in this case five steps). Further, we find that people are scope sensitive in a contingent valuation setting even under DA. Notably, this scope sensitivity can further be enhanced using SUA but not using CUA.

## 9 | GENERAL DISCUSSION

We tested whether Unit Asking makes people scope sensitive as claimed in previous research. We found that Unit Asking only gives a

one-off boost to WTD judgments, independent of scope. In other words, participants donated more under Unit Asking as opposed to DA; however, this increase was independent of the number of individuals affected and, therefore, does not seem to reflect genuine scope sensitivity.

We also introduced a new variant of Unit Asking, which we call SUA. SUA extends CUA by asking a sequence of questions scaling up with scope. We found evidence in three out of four studies that SUA increased WTD judgments over CUA. In addition, when pooling across all studies, we found overall evidence that SUA increased WTD judgments over CUA. However, this increase also seems to come only as a limited series of one-off boosts rather than covarying with scope.

We further investigated contingent valuation judgments, where we found the inverse pattern in comparison with WTD judgments. People were scope sensitive under all of the asking techniques, but CUA and SUA did not strongly increase judgments in comparison with DA. Table 2 gives an overview of the results across all five studies.

Finally, we advanced the methodology for analyzing WTD judgments. We showed that the oft-used method of winsorizing and using

**TABLE 2** Summary of results across all four studies.

Study	Mean (median) DA	Mean (median) CUA	Mean (median) SUA	Evidence for SUA over CUA	Evidence for scope sensitivity CUA	Evidence for scope sensitivity SUA
Donation judgments						
Study 1a	\$27.36 (\$10)	\$81.13 (\$20)	\$177.58 (\$50)	+++	N/A	N/A
Study 1b	\$42.14 (\$20)	\$72.70 (\$25)	\$231.01 (\$50)	+++	N/A	N/A
Study 2	\$85.92 (\$10)	\$136.74 (\$25)	\$257.01 (\$50)	+++	–	–
Study 3	N/A	\$86.36 (\$30)	\$97.27 (\$50)	–	--	–
Contingent valuation judgments						
Study 4 (Scope 10)	\$479.34 (\$300)	\$637.97 (\$500)	\$627.77 (\$500)	–	+++	+++*
Study 4 (Scope 50)	\$1853.51 (\$1000)	\$1697.73 (\$1000)	\$2059.93 (\$1375)			

Note: “+++” denotes strong evidence for  $\mathcal{H}_1$ , “++” denotes moderate evidence for  $\mathcal{H}_1$ , “+” denotes weak evidence for  $\mathcal{H}_1$ , “–” denotes weak evidence for  $\mathcal{H}_0$ , “--” denotes moderate evidence for  $\mathcal{H}_0$ , and “---” denotes strong evidence for  $\mathcal{H}_0$ . Medians and means here reported including donations of 0. SUA uses the number of steps constant condition for scaling up the scope. For Study 2, SUA refers to the condition, where the maximum scope was not known for comparability. As only for Study 4, the judgments differed strongly based on scope; we distinguish between Scope 10 and Scope 50 for this study. \*Unlike CUA, SUA further increased scope sensitivity above DA.

Abbreviations: DA, Direct Asking; CUA, Classical Unit Asking; SUA, Sequential Unit Asking.

conventional *t*-tests is problematic, as even after winsorizing the inference is not robust to outliers. Instead, we used a Bayesian model with lognormal likelihoods that can directly accommodate skew. We share the analysis code in our OSF project so that other researchers might use the methodology when analyzing WTD judgments or other judgments with similar distributional properties (i.e., extreme positive skew and zeros are reasonably excluded).

## 9.1 | Why is it Difficult to Create Scope Sensitivity in WTD Judgments and Easier in Contingent Valuation?

A possible explanation for the difficulty of making people scope sensitive in WTD judgments using the different Unit Asking manipulations is based on mental accounting. In line with research on this topic (Sussman et al., 2015; Thaler, 1985, 1999), participants might have a fixed budget of how much they are willing to give to charity. This allocated money might already be exploited on relatively small numbers of children; therefore, participants do not usually increase their WTD judgments anymore for larger scopes in the Unit Asking condition. This is also reflected in comments that we got from participants in Study 2, where we included an open feedback box. For example, one participant indicated “I have a certain amount total I'm willing to give (\$15), so after that is reached I'm not willing to give anymore.” and another participant said “I have limited funds so I can only donate so much regardless of the number in need.” Versions of these comments were echoed by a considerable proportion of participants. Even though we tried to mitigate the problem by reducing the maximum scope in Study 3, it is still possible that both studies had reached the donation ceiling for most participants on both scopes, thus diminishing scope sensitivity.

In contrast, the contingent valuation setting removes this possibility: participants are asked how much money is *needed* to buy

Christmas presents. This explains why participants are more readily scope sensitive even without using CUA or SUA. This baseline scope sensitivity can be further enhanced by SUA. However, we only observe a weak form of scope sensitivity, where participants donate more for more children, but this number does not increase proportionally with the scope, a finding that is qualitatively in line with previous literature on scope insensitivity (Dickert et al., 2012, 2015).

Overall, we conclude that when (hypothetical) personal funds are at stake, participants are reluctant to even show weak scope sensitivity, and only strong manipulations such as repeated asking will induce increases in donations. On the other hand, people readily show scope sensitivity for contingent valuation judgments.

## 9.2 | Future directions

In this manuscript, we have focused solely on hypothetical WTD judgments. A standard suggestion for future research would therefore be to include incentive compatible studies. We do not, however, perceive these as a fruitful avenue for future research in this area for three reasons. First, in the current studies, our observations of different results between WTD judgments and contingent valuation judgments, coupled with participants' references to budgetary constraints in open-text feedback, suggest that our participants did take the hypothetical judgments seriously and were restricted by real-world budgets. Second, the long ongoing debate about how much financial incentives change participant behavior in the social and behavioral sciences (Camerer & Hogarth, 1999; Hertwig & Ortmann, 2001) mostly concludes that “the effects of incentives are mixed and complicated” (Camerer & Hogarth, 1999, p. 1). While research has found that reliance on hypothetical donations increases mean donations (Bekkers, 2017), to our knowledge, there is no evidence that hypotheticality influences differences between conditions in WTD judgments. Finally, incentive

compatible studies typically provide an endowment and ask participants to donate a proportion of this (e.g., Schoenegger & Costa-Gomes, 2022; Small et al., 2007), which is a considerably different task from the majority of real-life donation decisions. Given that our participants appeared to take the WTD task seriously (as evidenced above), we would see such an incentive compatible task as more different from real donation decisions than our hypothetical tasks, especially as it induces an upper bound on how much participants can donate (the endowment), which would make it more difficult to study scope sensitivity.

While we do not see the hypotheticality of the current judgments as limiting the generalizability of the current results, we suggest that future research should seek to generalize them beyond an experimental setting. For real donation decisions, there may be a trade-off between the “boost” provided by SUA and the need to maintain potential donors' attention and reduce the number of questions asked. This question could not be answered within any lab/experimental context. Regardless of whether incentives are offered or not, experimental participants have a reasonable expectation of answering a series of questions. To determine how to maintain potential donors' interest and lever the benefits of SUA in the real-world, a naturalistic field study would be required. Such a study would enable stronger conclusions as to the effectiveness of SUA “in increasing charitable donations in the wild.”

Further, in line with previous studies on Unit Asking (Hsee et al., 2013; Karlsson et al., 2020), we only include a picture of one needy child and do not increase the number of children in the picture as scope increases. The fact that the Unit Asking method increases average donations provides evidence that participants are able to scale their concern with the number of children, at least in an ordinal way, even when visualizations of the scope are not provided. However, adding visualizations to represent additional children may help induce scope sensitivity in combination with (Sequential) Unit Asking. Identifying appropriate visualizations might therefore be one avenue for future research to further increase scope sensitivity (see, e.g., this educational video for visualization/animation techniques: <https://www.youtube.com/watch?v=LEENEFaVUZU>).

Finally, future work may also explore different sequences of steps. Study 2 showed that the effect of SUA levels off when including too many steps. It would be interesting to further investigate at which point the effect of SUA begins to level off (although leveling off would likely depend on a number of contextual factors, such as the maximum scope). In addition to changing the number of steps, one could also change the ordering of steps or include steps in random order to further investigate how the effects are shaped by the step order.

### 9.3 | Conclusion

This paper showed the difficulty in inducing scope sensitivity in WTD judgments. Contrary to previous claims, Unit Asking does not increase scope sensitivity in separate evaluation but instead gives a one-off

boost to WTD judgments. However, our findings also suggest that this one-off boost is increased by asking additional stepwise questions, a technique we term SUA.

### ACKNOWLEDGMENTS

We thank Henrik Singmann for advice with the Bayesian analysis in BRMS and Vanessa Cheung for feedback on a previous version of the manuscript. We conducted two additional pilots that we considered less relevant for the main insights in this paper, which are reported in the Appendix. Data, code, materials, and preregistration are available online (<https://osf.io/qjuhb/>). Note that we improved the analysis throughout the project. To keep consistency, we present all the studies here with the analysis technique preregistered for Study 3. An alternative write-up with analyses exactly as preregistered is available in the OSF project. The conclusions are the same in terms of the Bayes factor categories unless stated otherwise.

### DATA AVAILABILITY STATEMENT

The data that support the findings, analysis code, and preregistration for this study are openly available in Open Science Framework at <https://osf.io/qjuhb/>.

### ORCID

Maximilian Maier  <https://orcid.org/0000-0002-9873-6096>

Adam J. L. Harris  <https://orcid.org/0000-0002-9041-6290>

### REFERENCES

- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118(1), 73–106. <https://doi.org/10.1162/00335530360535153>
- Baron, J., & Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*, 2(2), 107. <https://doi.org/10.1037/1076-898X.2.2.107>
- Bekkers, R. (2017). Words and deeds of generosity: Are decisions about real and hypothetical money really different? [https://renebekkers.files.wordpress.com/2015/12/15\\_12\\_01\\_words\\_and\\_deeds.pdf](https://renebekkers.files.wordpress.com/2015/12/15_12_01_words_and_deeds.pdf)
- Bürkner, P.-C. (2017). BRMS: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package BRMS. *The R Journal*, 10(1), 395–411.
- Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19(1), 7–42. <https://doi.org/10.1023/A:1007850605129>
- Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, 100(1), 7–42. <https://doi.org/10.1023/A:1007850605129>
- Desvousges, W. H., Johnson, F. R., Dunford, R. W., Hudson, S. P., Wilson, K. N., & Boyle, K. J. (1993). Measuring natural resource damages with contingent valuation: tests of validity and reliability. In Baltagi, B. H., & Moscore, F. (Eds.), *Contributions to economic analysis*, Vol. 220, pp. 91–164. <https://doi.org/10.1016/B978-0-444-81469-2.50009-2>
- Dickert, S., Vstfjll, D., Kleber, J., & Slovic, P. (2012). Valuations of human lives: Normative expectations and psychological mechanisms of



- (IR) rationality. *Synthese*, 189(1), 95–105. <https://doi.org/10.1007/s11229-012-0137-4>
- Dickert, S., Vstfjll, D., Kleber, J., & Slovic, P. (2015). Scope insensitivity: The limits of intuitive valuation of human lives in public policy. *Journal of Applied Research in Memory and Cognition*, 4(3), 248–255. <https://doi.org/10.1016/j.jarmac.2014.09.002>
- Etz, A., & Wagenmakers, E.-J. (2017). J. B. S. Haldane's contribution to the Bayes factor hypothesis test. *Statistical Science*, 32, 313–329. <https://doi.org/10.1214/16-STS599>
- Fechner, G. T. (1860). *Elemente der psychophysik*. Vol. 2. Breitkopf u. Härtel.
- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., & Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, 81, 80–97. <https://doi.org/10.1016/j.jmp.2017.09.005>
- Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). bridgesampling: An R package for estimating normalizing constants. <https://doi.org/10.48550/arXiv.1710.08162>
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24(3), 383–403. <https://doi.org/10.1017/S0140525X01004149>
- Hsee, C. K., Zhang, J., Lu, Z. Y., & Xu, F. (2013). Unit asking: A method to boost donations and beyond. *Psychological Science*, 24(9), 1801–1808. <https://doi.org/10.1177/0956797613482947>
- Jeffreys, H. (1939). *Theory of probability* (1st ed.). Oxford, UK: Oxford University Press.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22(1), 57–70. [https://doi.org/10.1016/0095-0696\(92\)90019-S](https://doi.org/10.1016/0095-0696(92)90019-S)
- Kahneman, D., Ritov, I., Schkade, D., Sherman, S. J., & Varian, H. R. (1999). Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. *Journal of Risk and Uncertainty*, 19, 203–242. <https://doi.org/10.1023/A:1007835629236>
- Karlsson, H., Hellström, S., Moche, H., & Vstfjll, D. (2020). Unit Asking—A method for increasing donations: A replication and extension. *Judgment and Decision Making*, 15(6), 989–993. <https://doi.org/10.1017/S1930297500008184>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- Kloke, J. D., & McKean, J. W. (2012). RFIT: Rank-based estimation for linear models. *Journal of R*, 4(2), 57.
- Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*: Cambridge University Press.
- Lopes, A. F., & Kipperberg, G. (2020). Diagnosing insensitivity to scope in contingent valuation. *Environmental and Resource Economics*, 77(1), 191–216. <https://doi.org/10.1007/s10640-020-00470-9>
- Luisetti, T., Bateman, I. J., & Turner, R. K. (2011). Testing the fundamental assumption of choice experiments: Are values absolute or relative? *Land Economics*, 87(2), 284–296.
- Marcinkiewicz, M. (2016). Increasing charitable donations: The limits of the unit asking effect. Unpublished Master Thesis.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.
- Mill, J. S. (1879). *Utilitarianism*: Fraser's Magazine.
- Morey, R. D., Rouder, J. N., Jamil, T., & Morey, M. R. D. (2015). Package 'bayesfactor'. Accessed 1006 15. <https://cran.r-project.org/web/packages/BayesFactor/index.html>
- Rouder, J. N., & Morey, R. D. (2019). Teaching Bayes theorem: Strength of evidence as predictive accuracy. *The American Statistician*, 73(2), 186–190. <https://doi.org/10.1080/00031305.2017.1341334>
- Schoenegger, P., & Costa-Gomes, M. (2022). Sure-thing vs. probabilistic charitable giving: Experimental evidence on the role of individual differences in risky and ambiguous charitable decision-making. *PLoS One*, 17(9), e0273971. <https://doi.org/10.1371/journal.pone.0273971>
- Simmons, J. (2013). A new way to increase charitable donations: Does it replicate? <http://datacolada.org/2013/10/02/3-a-new-way-to-increase-charitable-donations-does-it-replicate/>
- Slovic, P., & Vstfjll, D. (2010). Affect, moral intuition, and risk. *Psychological Inquiry*, 21(4), 387–398. <https://doi.org/10.1080/1047840X.2010.521119>
- Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143–153. <https://doi.org/10.1016/j.obhdp.2006.01.005>
- Stefan, A. M., Evans, N. J., & Wagenmakers, E. J. (2022). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*, 27(2), 177–197. <https://doi.org/10.1037/met0000354>
- Sussman, A. B., Sharma, E., & Alter, A. L. (2015). Framing charitable donations as exceptional expenses increases giving. *Journal of Experimental Psychology: Applied*, 21(2), 130. <https://doi.org/10.1037/xap0000047>
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199–214. <https://doi.org/10.1287/mksc.4.3.199>
- Thaler, R. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3), 183–206.
- Thomas, K. J., Hamilton, B. C., & Loughran, T. A. (2018). Testing the transitivity of reported risk perceptions: Evidence of coherent arbitrariness. *Criminology*, 56(1), 59–86. <https://doi.org/10.1111/1745-9125.12154>
- Västfjäll, D., & Slovic, P. (2020). A psychological perspective on charitable giving and monetary donations: The role of affect. In Zaleskiewicz, T., & Traczyk, J. (Eds.), *Psychological perspectives on financial decision making* (pp. 331–345). Springer. [https://doi.org/10.1007/978-3-030-45500-2\\_14](https://doi.org/10.1007/978-3-030-45500-2_14)
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistical Computing*, 27, 1413–1432.
- Veisten, K., Hoen, H. F., Navrud, S., & Strand, J. (2004). Scope insensitivity in contingent valuation of complex environmental amenities. *Journal of Environmental Management*, 73(4), 317–331. <https://doi.org/10.1016/j.jenvman.2004.07.008>
- Wagenmakers, E.-J. (2020). Bayesian thinking for toddlers.
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25(3), 169–176. <https://doi.org/10.1177/0963721416643289>
- Wrinch, D., & Jeffreys, H. (1921). On certain fundamental principles of scientific inquiry. *Philosophical Magazine*, 42, 369–390. <https://doi.org/10.1080/14786442108633773>

**How to cite this article:** Maier, M., Caviola, L., Schubert, S., & Harris, A. J. L. (2023). Investigating (sequential) unit asking: An unsuccessful quest for scope sensitivity in willingness to donate judgments. *Journal of Behavioral Decision Making*, 36(4), e2335. <https://doi.org/10.1002/bdm.2335>

## APPENDIX A: PILOT 2—BUDGET CONSTRAINS AS EXPLANATION FOR THE EFFECTIVENESS OF SUA OVER CUA

In pilot 2, we aimed to investigate whether a possible mechanism explaining the effectiveness of SUA was that people disengage when they need to increase the donation too much in one step. In other words, they might donate beyond budget constraints. We manipulated this by starting out from a small donation of \$0.1 for one child resulting in a small donation even after multiplying with 100 versus a larger donation of \$10 where people should disengage when directly jumping to a consistent donation of \$1000 for the full scope.

### A.1 | Methods

#### A.1.1 | Experimental conditions

This study employed a 2 (technique: SUA vs. CUA)  $\times$  2 (costs: small vs. large) between-subjects design. We manipulated the costs required to donate to help one child. The vignette was similar to Study 1, with the modification that this time participants were asked to donate to help treat parasitic worms, with treatment costing either large (\$10) or small (\$0.1). One group was told that 10 cents are needed to help one child that is sick of parasitic worms, while the other group that \$10 are needed to cure this disease. In addition, we varied SUA versus CUA. We hypothesized that both interventions should work similarly when only 10 cents were required to help one child but that SUA would work remarkably better when \$10 was required to help one child. In other words, there would be an interaction between Asking Type (SUA vs. CUA) and the costs of saving one child. We also randomized whether money was displayed in cents or in dollars to check for an effect of the currency unit. In addition, for our analysis, we rescaled the group giving only \$0.1 by multiplying it by 1000 so that the interaction effect could be analyzed.

The vignette was similar to Study 1, with the modification that this time participants were asked to donate to help treat parasitic worms, with treatment costing either large (\$10) or small (\$0.1). In addition, all values were presented on the cents scale and the dollar scale since we had suspected that participants are primed to donate more when the currency unit is dollar compared with cents. To check this idea, half of the participants saw all units in dollars and the other half in cents.

#### A.1.2 | Analyses

We used a Bayesian ANOVA in the package `BRMS` as in the previous study. Because we cannot meaningfully analyze scope insensitivity for participants that donated 0 (as they would donate 0 for any scope under scope consistency as well as scope inconsistency), we excluded these participants. We excluded unbelievably large WTD

judgments of more than \$10,000. As our data were (after excluding the zeros) expected to be positive and quite skewed, we used a log-normal likelihood rather than the more common Gaussian likelihood. We used a prior of normal (3,1) on the intercept, normal (0, 0.4) on the scope (as we expected scope insensitivity), and normal (0.4, 0.4) on the main effect of Asking Type, normal (0.25, 0.25) on the interaction, and normal (1, 0.5) on  $\sigma$ . The reason why the priors are increasingly more narrow is that given the lognormal likelihood, the prediction on the linear scale corresponds to the exponent of the marginal mean. Therefore, using similar priors on main effects in interactions would result in the prediction of an extremely large interaction effect on the linear scale. We tested for the presence of the hypothesized interactions by comparing models that include the interaction to models that do not include the interaction using Bayes factors. We also conducted a nonparametric analysis using robust ANOVA in the `Rfit` package.

#### A.1.3 | Participants

The study was run on January 23, 2021, targeting American participants through Positly. We paid participants \$0.31 to participate in a 2.5-min study. Two participants were excluded because they indicated extremely high or negative age. The mean participant age was 443.12 ( $sd = 14.14$ ). Ninety-seven participants were female, 103 were male, and 1 did not indicate their gender. This study was approved by the University of Oxford Central University Research Ethics Committee (reference number R56657/RE001).

#### A.1.4 | Results and discussion

A Bayesian contingency table analysis indicated that the intervention did not influence the share of participants donating ( $BF_{10} = 0.19$ ). Therefore, we proceeded to the main analysis, excluding 25 participants that donated 0 and 8 participants that donated more than 10,000. We find no evidence for a main effect of SUA versus CUA ( $BF_{10} = 1.11$ ) but evidence that people donate more when currencies are presented in terms of dollars rather than cents ( $BF_{10} = 8.67$ ). We find weak evidence against the predicted interaction of Asking Type with scope ( $BF_{10} = 0.55$ ). This is also corroborated by the robust ANOVA which finds no evidence for a main effect of Asking Type,  $F(1,163) = 0.22$ ,  $p = .637$ , a main effect of cents versus dollars,  $F(1,163) = 4.02$ ,  $p = .047$ , and no evidence for an interaction effect of starting value and Asking Type,  $F(1,163) = 0.010$ ,  $p = .752$ .

To sum up, this study neither found a main effect of SUA nor the predicted interaction. We believe a crucial flaw of this study was that telling participants how much would be needed to save one child beforehand undermined the point of SUA or CUA because participants could already multiply without using any specific asking technique. This is also why we did not include it in the main part of the manuscript.

## APPENDIX B: PILOT 3—PRIMING SCOPE SENSITIVITY

### B.1 | Methods

The aims of this study were twofold. First is to do a more fine-grained analysis of the SUA technique. We varied the number of steps and investigated how that affects the amount that participants donate after the final step. In addition, we varied the difference between steps to see how this impacts the WTD judgments. Second, we tested whether, after being consistent for several small steps, people are primed to be consistent, thus making them more consistent in a later larger step.

1. Consistency priming: 1, 2, 3, 4, 5, 100 [6 steps]
2. Equal spacing: 1, 20, 40, 60, 80, 100 [6 steps]
3. Max steps: 1, 2, 3, 4, 5, 20, 40, 60, 80, 100 [10 steps]

Our first hypothesis was the people would give the most in the Max steps condition because there is some increase with each step. Our second hypothesis was that people give more in the consistency priming condition than in the equal spacing condition due to the priming idea discussed above. Our third hypothesis was that people would give more in the next-to-last step in the equal spacing condition (where there are 80 beneficiaries) than in the consistency priming condition (where there are five beneficiaries). This would show that conditional on the number of steps being fixed (and the size of all steps being the same), the participants will donate more the greater the size of each step is; in other words, they are at least somewhat scope sensitive.

#### B.1.1 | Analyses

We used a Bayesian ANOVA in the package `BRMS` as in the previous study. Because we cannot meaningfully analyze scope insensitivity for participants that donated 0 (as they would donate 0 for any scope under scope consistency as well as scope inconsistency), we excluded these participants. As preregistered, we excluded unbelievably large WTD judgments of more than \$10,000. As our data were (after excluding the zeros) expected to be positive and quite skewed, we used a lognormal likelihood rather than the more common Gaussian likelihood. We used a prior of normal (3,1) on the intercept, normal (0, 0.4) on the scope (as we expected scope insensitivity), and normal (0.4, 0.4) on the main effect of Asking Type, normal (1, 0.5) on  $\sigma$ . The reason why the priors are increasingly more narrow is that given the lognormal likelihood, the prediction on the linear scale corresponds to the exponent of the marginal mean. Therefore, using similar priors on intercepts and slopes would result in

the prediction of an extremely large difference between the groups. We tested for the presence of the hypothesized mean differences by comparing models that include the interaction to models that do not include the interaction using Bayes factors. In addition, we used nonparametric Wilcoxon tests to check the robustness of our conclusions.

#### B.1.2 | Participants

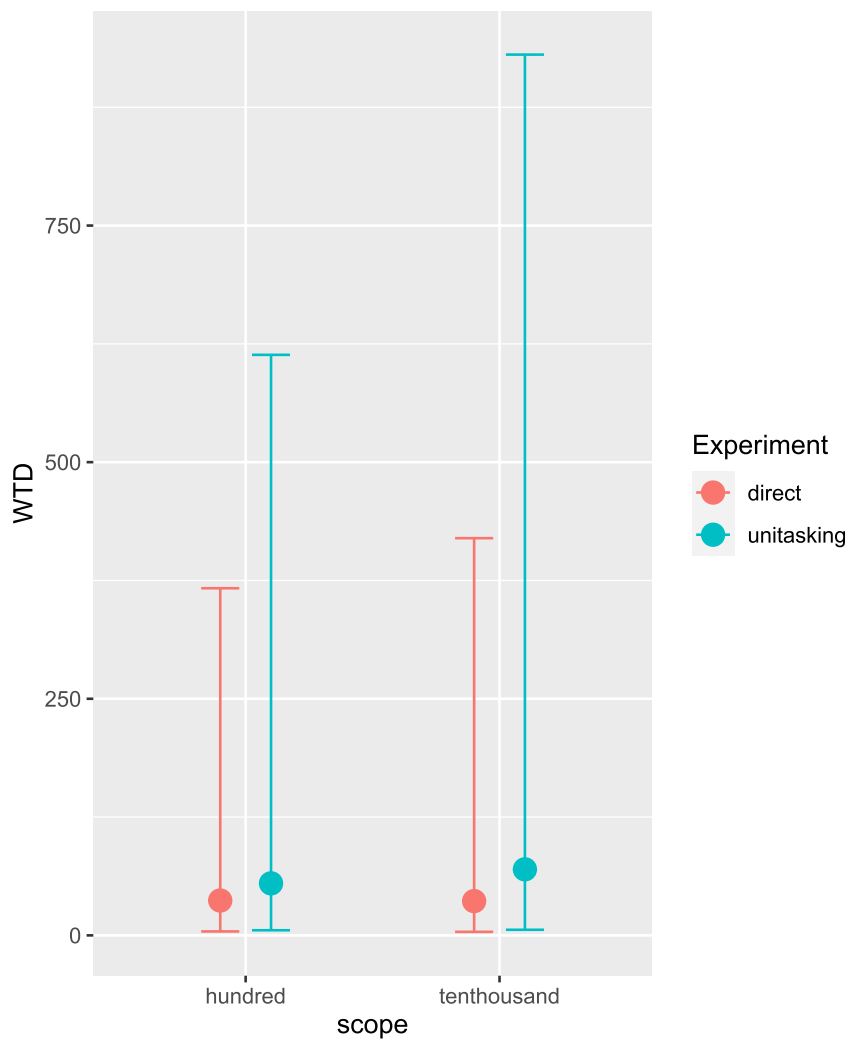
We recruited American participants through Positly. The study was run on February 18, 2021. Participants received a payment of \$0.36 to participate in a 3-min study. Initially, 201 participants signed up; we excluded two for indicating implausible age. In addition, we excluded one participant who donated more than \$10,000. We also excluded two participants who indicated impossible ages. This left us with a final sample of 177 participants. Mean age was 43.39 ( $sd = 12.06$ ). Ninety-one participants were female, and 108 were male. This study received ethical approval by Harvard University's ethics review board.

#### B.1.3 | Results and discussion

As all conditions in this study were SUA conditions, we directly proceeded with excluding WTD judgments of 0 (21 participants) and more than 10,000 (1 participant). We only found weak evidence that more steps lead to higher WTD judgments ( $bf_{10} = 2.12$ ). Hypothesis two was also not supported with only weak evidence that consistency priming improved performance over equal spacing ( $bf_{10} = 1.97$ ). Finally, we also found strong evidence that participants donated more for 80 in the equal spacing condition than for 5 in the consistency priming condition ( $bf_{10} = 23.83$ ). This is also corroborated by non-parametric Wilcoxon test (Hypothesis one:  $W = 1585.5$ ,  $p = .1276$ ; Hypothesis two:  $W = 2800.5$ ,  $p = .095$ ; Hypothesis three:  $W = 1307.5$ ,  $p = .005$ ).

## APPENDIX C: PRIOR PREDICTIVES FOR BRMS MODEL

When sampling only from priors, we obtain predictions for the marginal means as in Figure C.1. In other words, we would estimate a donation of 37 for the DA condition and a donation of 56 for the unit asking condition on a low scope. Because of the hypothesized interaction of Asking Type and scope under  $H_1$ , we predicted a donation of 71 for unit asking and scope 10,000. The skew of the credible intervals is a feature of the lognormal likelihood, which helps accommodate strong positive outliers.



**FIGURE C.1** Prior predictives for the lognormal model.

**APPENDIX D: SHARE OF PARTICIPANTS DONATING FOR DA, CUA, AND SUA CONDITIONS IN THE DIFFERENT STUDIES**

**TABLE D1** Median WTD judgments for the seven conditions of Study 2.

Condition	Scope	Median	% Donating
DA	100	\$15	36.59
CUA	100	\$40.0	17.28
SUA	100	\$79.0	14.81
DA	10,000	\$30.0	30.00
CUA	10,000	\$50.0	23.75
SUA	10,000	\$75.0	15.19
SUA <sup>a</sup>	10,000	\$50.0	27.85

Abbreviations: DA, Direct Asking; CUA, Classical Unit Asking; SUA, Sequential Unit Asking; WTD, willingness to donate.

<sup>a</sup>SUA scaling up with increase per step constant.

**TABLE D2** Median WTD judgments for the seven conditions of Study 3.

Condition	Scope	Median	% Donating
CUA	10	\$50.0	9.76
SUA	10	\$50.0	15.48
SUA	10 & scope known	\$50.0	11.39
CUA	50	\$40.0	8.43
SUA	50	\$50.0	14.63
SUA	50 & scope known	\$50.0	10.84
SUA <sup>a</sup>	50	\$75.0	14.81

Abbreviations: CUA, Classical Unit Asking; SUA, Sequential Unit Asking.

<sup>a</sup>SUA scaling up with increase per step constant.

## APPENDIX E: STEP-BY-STEP GUIDE TO THE BAYESIAN ANALYSIS

This Appendix contains a step-by-step guide to the Bayesian analysis executed in this paper, including R code. All analyses use R version 4.1.2, and the packages `brms` version 2.16.3, `rstan` version 2.21.3, and `BayesFactor` version 0.9.12. This Appendix is merely intended as a conceptual explanation for the most important parts of the R code; for a more general accessible introductions to Bayesian statistics, see Wagenmakers et al. (2016) and Wagenmakers (2020).

### E.1 | Step 1: Testing whether proportion of participants donating in the first place is affected

After reading in the data (and excluding participants that failed attention checks), we use the `contingencyTableBF` function from the `BayesFactor` package to check whether the intervention influenced the share of participants donating in the first place.

```
#Construct Matrix of Counts
counts <- c(sum(na.omit(subset(data, Experiment..
Pilot. == "Control")$WTD) > 0),
sum(na.omit(subset(data, Experiment..Pilot. ==
"Control")$WTD) == 0),
sum(na.omit(subset(data, Experiment..Pilot. ==
"Intervention1")$WTD) > 0),
sum(na.omit(subset(data, Experiment..Pilot. ==
"Intervention1")$WTD) == 0),
sum(na.omit(subset(data, Experiment..Pilot. ==
"Intervention2")$WTD) > 0),
sum(na.omit(subset(data, Experiment..Pilot. ==
"Intervention2")$WTD) == 0))
counts <- matrix(counts, nrow = 2)
#inspect proportion of people donating 0
counts[2,]/colSums(counts)
#Check whether it depends on the manipulation
BayesFactor::contingencyTableBF(counts, sample-
Type = "jointMulti")
```

### E.2 | Step 2: Fitting the models

Once we established that the proportion of participants donating in the first place is not affected, we can go ahead and fit the models. In order to test for the presence of a difference between groups, we specify one model that assumes a difference between groups ( $\mathcal{H}_1$ ) to one model that does not assume a difference ( $\mathcal{H}_0$ ).

```
#H1
data.brms1 = brm(bf(WTD ~ Experiment..Pilot., cen-
ter = FALSE),
data = data, iter = 10000, warmup = 2000, chains = 4,
cores = 12, thin = 2, refresh = 0, family = lognormal
()),
prior = c(prior(normal(0.4, 0.4), class = "b"),
prior(normal(3, 1), class = "b", coef =
"Intercept")),
prior(normal(1, 0.5), class = "sigma")
), save_all_pars = T)
#H0
data.brms0 = brm(bf(WTD ~ 1, center = FALSE),
data = data, iter = 10000, warmup = 2000, chains = 4,
cores = 12, thin = 2, refresh = 0, family = lognormal
()),
prior = c(prior(normal(3, 1), class = "b", coef =
"Intercept"), prior(normal(1, 0.5), class = "sigma")
), save_all_pars = T)
```

We can see that this code specifies two competing models: the  $\mathcal{H}_1$  includes an effect of the intervention ( $WTD \sim \text{Experiment}..Pilot.$ ), whereas the  $\mathcal{H}_0$  model does not ( $WTD \sim \text{Experiment}..Pilot.$ ). Further, we specify the number of iterations and warmup for the MCMC sampling (details on MCMC sampling and checking of model convergence are beyond the scope of this guide, see, for example, McElreath, 2020) and prior distributions. The priors here were based on pilot data. There is ample guidance available on different approaches to specifying prior distributions, and we refer the reader to the literature (e.g., Stefan et al., 2022). Finally, we specified `family = lognormal()`. This indicates that we are relying on a lognormal likelihood rather than a normal likelihood.

### E.2.1 | Step 3: Comparing the models

Finally, we compare the two models under consideration using Bayes factors:

```
bayes_factor(data.brms1, data.brms0)
```

This code will use bridge sampling to estimate the Bayes factors (more details in Gronau et al., 2017). The Bayes factor compares the likelihood of the data under the two competing models. Here we calculate the Bayes factor in favor of the alternative (as `data.brms1` is listed first in the brackets). As a rule of thumb, Bayes factors between 1 and 3 are considered weak evidence, Bayes factors between 3 and 10 are considered moderate evidence, and Bayes factors larger 10 are considered strong evidence (Lee & Wagenmakers, 2013). When the evidence for the null is considered, the Bayes factor is simply inverted ( $BF_{10} = 1/BF_{01}$ ).