

König-Kersting, Christian

Working Paper

On the Robustness of Social Norm Elicitation

Working Papers in Economics and Statistics, No. 2021-02

Provided in Cooperation with:

Institute of Public Finance, University of Innsbruck

Suggested Citation: König-Kersting, Christian (2021) : On the Robustness of Social Norm Elicitation, Working Papers in Economics and Statistics, No. 2021-02, University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon), Innsbruck

This Version is available at:

<https://hdl.handle.net/10419/238260>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



On the Robustness of Social Norm Elicitation

Christian König-Kersting

Working Papers in Economics and Statistics

2021-02



University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
research platform "Empirical and Experimental Economics"
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 71022
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
<https://www.uibk.ac.at/eeecon/wopec/>

For a list of recent papers see the backpages of this paper.

On the Robustness of Social Norm Elicitation

Christian König-Kersting^{1*}

¹*University of Innsbruck, Austria*

6 January 2021

We study the robustness of Krupka and Weber's method (2013) for eliciting social norms. In two experiments with more than 1200 participants, we find that participants' response patterns are invariant to differences in the salience of the monetarily incentivized coordination aspect. We further demonstrate that asking participants for their personal first and second order beliefs without monetary incentives results in qualitatively identical responses. In addition, we observe that participants give sensible responses whether or not they understand the task or their monetary incentives. Overall, Krupka and Weber's method produces remarkably robust response patterns.

JEL: C72, C90, D90

Keywords: social norms, incentives, beliefs, task comprehension, robustness.

* Christian König-Kersting: Department of Banking and Finance, University of Innsbruck, Universitätsstraße 15, 6020 Innsbruck, Austria, Phone: +43 512 507 73007, email: christian.koenig@uibk.ac.at. I thank Christoph Huber and Matthias Stefan for helpful comments.

1 Introduction

We report on two online experiments with more than 1200 participants designed to study the robustness of Krupka and Weber's (2013, KW henceforth) method for social norm elicitation. Specifically, we test how the salience of the coordination game aspect and the monetary incentives affect participants' responses. We also compare responses elicited using their method to non-incentivized first- and second-order beliefs. Finally, we assess how attention to the instructions and task understanding affect participants' responses.

Krupka and Weber's method has been widely used to elicit social norm perception in economic experiments. The main feature of the elicitation method is that it incentivizes participants to form a belief about what the modal response of all study participants is to a given question. For example, in the experiment of the original article of KW, participants are incentivized to correctly guess the modal social appropriateness rating of a series of hypothetical dictator game decisions.

There are substantial differences in the implementation of the method across different studies. Some implementations ask participants to state their personal believe, but incentivize them to report what they believe most others to believe (e.g., Krupka and Weber 2013, 2017; Erkut et al. 2015; Kimbrough and Vostroknutov 2016). As one's personal belief and one's belief about what most others think about an issue do not necessarily have to coincide,¹ this creates a dissonance or even conflict. Clearly, instructions which allow for multiple interpretations lead to more noisy responses and reduced data quality. There are studies in which participants are constantly reminded of their incentives (e.g., Barr et al. 2017; Erkut et al. 2015; Fallucchi and Nosenzo 2020; Zhang et al. 2018), and others, in which participants are never reminded of them after the initial instructions (e.g., Gächter et al. 2013, 2017; Abbink et al. 2017; Vesely and Klöckner 2017; Huber and Huber 2020).² If a dissonance exists between task description and incentivization, changing the salience of the incentives might affect participants' responses. In further studies, the potential conflict has been resolved by changing the wording of the instructions (e.g., Sass et al. 2018; Kölle et al. 2020; Heinicke et al. 2020). If participants'

¹ For example, one might hold the personal belief that allocation mechanisms should be procedurally fair, i.e., give everyone the *same chance* of obtaining the better of multiple outcomes. Yet, the same person might hold the belief that most others will favor allocation mechanisms that establish outcome fairness, i.e., give everyone the *same outcome*.

² Some studies adapted the wording, but do not fully resolve the dissonance between task and incentives. Notable examples are Chang et al. (2019) and D'Adda et al. (2016). There are also studies in which participants are not incentivized to select the same rating as most others, but are randomly matched in pairs for the purpose of payment determination (e.g., Barr et al. 2017).

responses are affected by details of the implementation, comparisons of results across studies becomes difficult, hindering the accumulation of knowledge.

Across four treatments, we systematically vary the salience of the conflict between task wording and financial incentives. We (i) replicate the original task wording of KW, (ii) increase the salience of the conflict between stated task and incentivization through constant reminders of the incentive structure on the decision pages, (iii) decrease the salience by never reminding participants of the incentives, and (iv) finally resolve the conflict by rephrasing the task. We do not find significant differences between the different variations in task wording, indicating that the elicitation method is largely robust to adaptations of the instructions.

The defining feature that sets KW's method for social norms elicitation apart from others, is its incorporation of monetary incentives. This not only makes the method attractive for economists, but ensures that revealing their true norm perception lies in participants' interest. A recent study shows that the response patterns do not substantially differ between using KW's elicitation method and a simpler approach of "just asking", without dedicated monetary incentives (Heinicke et al. 2020). Relatedly, Veselý (2015) do not find significant differences between incentivized and non-incentivized social appropriateness ratings for *ultimatum game* behavior. To formally test for differences between incentivized and non-incentivized response patterns in the context of *dictator* games, which are more prevalent among the studies using KW's method, we conduct two additional treatments in which we ask participants to state their first- and second-order beliefs about the appropriateness of the various dictator game allocations. That is, we ask participants, without monetarily incentivizing them, to state (v) what they personally believe to be socially appropriate, and (vi) what they think most people would consider socially appropriate behavior. We do not find responses in the incentivized KW elicitation procedure to differ significantly from stated first- and second-order beliefs.

After conducting these treatments in a first experiment, we realized that more than half of our participants displayed poor task understanding as revealed by a *post-task* questionnaire asking them to recall the task they had been given and the monetary incentives that had been put in place (if any).³ In response, we conducted a second experiment, again encompassing all six treatments. We added the two *post-task* questions as mandatory comprehension checks to

³ Note that response patterns in the main KW task are quite similar in all treatments, irrespective of them being monetarily incentivized or not. Thus, it is unlikely that the poor task understanding revealed by the post-task questionnaire is purely the result of the questionnaire not being incentivized and participants considering this part less important.

the *pre-task* instructions, enforcing correct responses. We find that these comprehension checks drastically improve post-task recall of task and incentives, but - strikingly - the results do not differ significantly between experiments 1 and 2. The response patterns also do not differ between those that appear to understand the task and those that do not. Throughout the paper we present data from all participants in both experiments and data from those participants that correctly answered the post-task questions side-by-side.

Both experiments and all treatments were specifically designed to represent the different variations found in past implementations of KW's method and tease out differences in behavior of respondents recruited from the same population. Across all treatments and both experiments, it has become clear that KW's method is remarkably robust and insensitive to variations in task wording, the salience of incentives, and task understanding. However, the responses elicited using their method also do not differ significantly from asking for participants' beliefs directly, without monetary incentives.

2 Design

The basic design of the experiment follows the “give”-framing of situation 1 in the first experiment reported in Krupka and Weber (2013). Participants in the experiment are asked to give appropriateness ratings for allocation decisions that a hypothetical dictator can take in a two-player dictator game. In the game, player A owns an endowment of \$10, while player B does not own anything (\$0). Player A can give any amount from \$0 to \$10 (in \$1 increments) to player B. For each of the 11 possible options (“give \$0” to “give \$10”) we ask participants in our experiment to rate how socially appropriate the action is perceived to be. Each option is presented separately and their order is randomized on the individual level. That is, we do not show a sorted choice list.

In our first *Baseline* treatment, participants are asked to “indicate whether [they] believe choosing that option is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate”. In addition, participants are reminded “that [they] will earn money [...] if [their] response to a randomly-selected question is the same as the most common response provided in today's session” (both quotes: Krupka and Weber, 2013, Online Appendix, p. 7). We use the same wording as in the original experiment. The wording creates an obvious dissonance: Participants are asked to respond according to their *individual* belief about the actual appropriateness, but are paid for indicating the *most common* response among all participants. Note that an individual's actual belief about the action's appropriateness

does not have to be in line with everyone else's belief. Take, for example, a rich participant who might believe it to be socially appropriate to only take 3 for themselves, but give 7 to the other participant, while being convinced that most people would consider an equal split allocation to be the socially appropriate choice. Krupka and Weber (2013) present participants with a series of situations, but only remind the participants of their financial response motives in the first. Similarly, in treatment *Baseline*, participants are only reminded of their financial incentives on the first appropriateness rating they provide. The reminder is not shown for subsequent ratings. *Baseline* conceptually replicates KW's original implementation.

We conduct three treatments aimed at identifying the sensitivity of the norm elicitation method to the two conflicting elements of the task description. In treatment *Always*, participants constantly, i.e., for each rating, see the reminder of their financial incentive to indicate the appropriateness rating that they believe most participants to select. This treatment is meant to reinforce the dissonance between stated task and financial incentives. In treatment *Never*, participants only learn about their financial incentive as part of the instructions, but are not reminded of it on the decision screen. Thus, this treatment deliberately attenuates the dissonance between task wording and monetary incentives. In treatment *No Conflict*, we modify the instructions to never allude to the participants' own appropriateness ratings. We explicitly ask them to indicate the rating that they believe *most people* to select, completely eliminating the conflict.

Furthermore, we conduct two treatments in which we remove monetary incentives and the coordination aspect. In *First*, we simply ask participants to give their individual appropriateness ratings. We avoid any references to other participants' views or additional payments beyond a fixed compensation for participating in the study. This treatment is free of conflict and asks for individual, first-order beliefs. In *Second*, we ask participants to state the appropriateness rating which they believe most participants in the session would give. Rather than attempting to elicit the social norm using a coordination device, we ask participants to reveal it directly. Table 1 shows an overview of all treatments.

Table 1: Treatment overview

		Task wording	
		Individual belief	Most people's beliefs
Payment	Incentivized	<i>Baseline</i>	
		<i>Always</i>	<i>Never</i>
	Non-incentivized	<i>First</i>	
		<i>Second</i>	

To stay true to previous implementations of the method in the literature, we did not include any comprehension questions in the instructions of the first experiment. Instead, we asked participants two questions about their task and their monetary incentives as part of a *post-task* questionnaire. Only 52%, respectively 44%, of our participants answered these questions correctly (see details in section 3). Acknowledging that the online setting may differ from a traditional laboratory environment in terms of participant dedication to the task, we opted to subsequently run a second experiment, involving an identical set of treatments. For this second experiment, we added the same two questions as comprehension questions to the *pre-task* instructions to force participants to engage more with the instructions and task description. Participants had to complete these questions correctly before continuing with the experiment. That is, in experiment 2, we asked the same questions in the *pre-task* and in the *post-task* questionnaire.

We recruited a total of 1228 participants from the USA on Amazon Mechanical Turk. Table A1 in Appendix A breaks down the number of participants by treatment and includes a small set of demographic variables. The average age of our participants is 37.5 years, 38.0% are female,⁴ and 2.2% report to have had some university-level education in economics. The experiment took approximately 6 minutes to complete and we paid a fixed amount of \$0.75 and an additional bonus of \$0.75 if a participant indicated the modal response in a randomly selected choice situation. In treatments *First* and *Second*, we only pay the fixed amount. On a per-hour basis, participants earned \$9.03 on average. The experiment was programmed with oTree (Chen et al. 2016).

⁴ One participant opted not to disclose their gender, 62.0% of our participants are male. For unknown reasons, there are more females in the second experiment than in the first: 35% vs. 41%, $p = 0.03$, two-sample test of proportions.

3 Results

We first present results pertaining to task understanding, as it is the main reason for conducting two experiments. In a second step, we establish the baseline norm response pattern. Then, we compare responses in the treatments designed to reinforce, attenuate, and eliminate the dissonance in the instructions to this baseline. Finally, we compare responses to KW's norm elicitation procedure to responses gathered by asking participants without using the element of coordination and monetary incentivization.

Task understanding. Responses to our comprehension checks in the first experiment reveal that, across all treatments, only 52% of our participants were able to correctly recall whether their task was to give appropriateness ratings based on a) their own personal belief or based on b) what they thought most people would believe. Similarly, across all treatments, just 44% of our participants were able to correctly identify whether their payment for the study a) depended on their ability to anticipate what most people believe, b) depended on their own personal beliefs, or c) was independent of their responses.

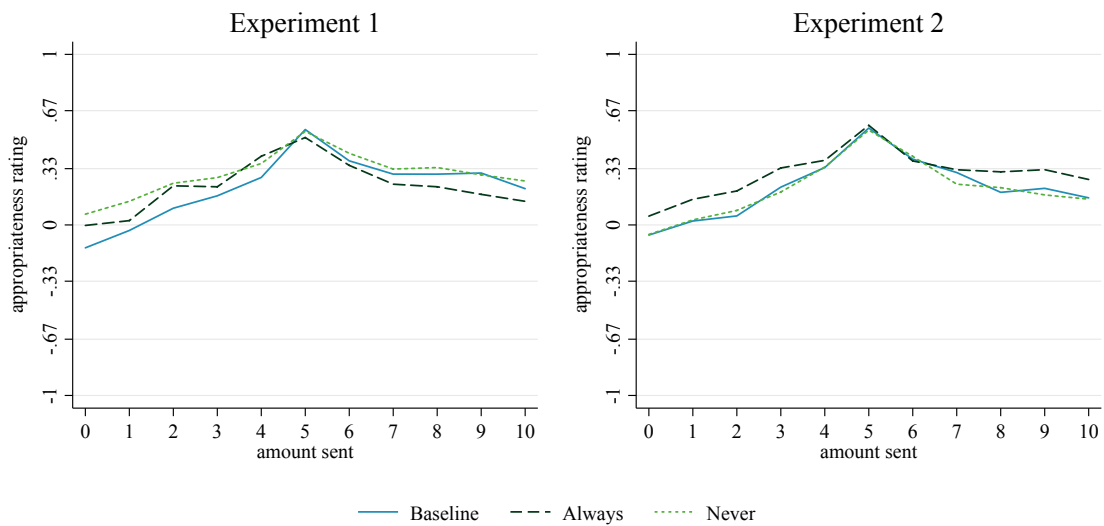
In the second experiment, in which participants had to answer the questions twice, once as part of the pre-task instructions and once in the post-task questionnaire, we observe an increase in the number of correct responses. In the post-task questionnaire of the second set of treatments, 69% correctly understood the task and 69% correctly understood the incentives (again across all treatments). Table 1 gives an overview of the shares of participants that answered the questions correctly in each treatment. Note that with and without pre-task comprehension checks, *No Conflict* shows the highest task understanding rates among all treatments based on KW's method. It appears that participants have a better grasp of their task when any potential dissonance between incentivization and actual task description is eliminated.

Table 1: Share of participants that correctly answered the comprehension checks

Treatment	Experiment 1	Experiment 2
<i>Task</i>		
Baseline	47.06%	62.38%
Always	51.89%	61.54%
Never	51.49%	67.65%
No Conflict	62.96%	79.21%
First	49.51%	68.32%
Second	51.00%	75.76%
	Ø 52.42%	Ø 69.08%
<i>Incentives</i>		
Baseline	52.94%	70.30%
Always	44.34%	61.54%
Never	43.56%	68.63%
No Conflict	38.39%	71.29%
First	43.69%	74.26%
Second	41.00%	68.69%
	Ø 44.03%	Ø 69.08%

Note: The table shows the share of participants that correctly answered the post-task comprehension questions. The top half shows data for the task question, the bottom half shows data for the incentives question.

Figure 1: Mean appropriateness ratings in *Baseline*, *Always*, and *Never*



Baseline. We observe the typical pattern of highest appropriateness ratings for 50/50 allocations between dictators and recipients, while lower and higher amounts transferred by the dictator are rated as less socially appropriate. The average ratings in our experiments lie in a relatively narrow band, ranging from -0.10 to 0.56 on the scale from -1 to 1.⁵ Figure 1 shows the average ratings for treatments *Baseline*, *Always*, and *Never*, separated by the first and second experiment. Differences between the two experiments are negligible (blue solid lines) for the *Baseline* treatment.

Reinforcing the dissonance. In a first step, we compare responses in *Always* to those in *Baseline*. The only difference between these treatments is how prominently participants are reminded of their monetary incentives. In *Baseline* they are only reminded when giving the first rating; in *Always*, the reminder is constantly shown on the decision screen. As Figure 1 suggests, there are no statistically significant differences in responses between the two treatments. This holds for both sets of treatments individually, as well as for the pooled data. Detailed test statistics of pairwise, two-sided Mann-Whitney-U tests for each allocation are reported in Table A2 in Appendix A.

Attenuating the dissonance. In a second step, we compare *Never* to *Baseline*. In *Never*, participants are informed about the monetary incentives as part of the instructions, but they are never reminded of these incentives when giving any of the appropriateness ratings. Compared to the *Baseline*, the salience of the dissonance between task description and incentivization is reduced. The response patterns are shown in Figure 1. We do not find systematic and significant differences between the responses in treatments *Baseline* and *Never*. This is true for both sets of treatments individually as well as the pooled data. Detailed test statistics are reported in Table A3 in Appendix A.

Eliminating the dissonance. In treatment *No Conflict* we modify the instructions to resolve the conflict between asking for personal opinions and paying for selecting the modal response. Participants are explicitly asked how most people would rate the allocations. Significant treatment differences (Mann-Whitney-U tests, $p < 0.01$) appear for small transfers (\$0 to \$3) in the first experiment and the pooled data, but not in the second experiment and for larger amounts (Figure 2, left panel). If we restrict the data to participants who correctly answered the question

⁵ Other online and laboratory experiments typically report ratings ranging from -0.80 to 0.90 (cf. Krupka and Weber 2013, Erkut et al. 2015, Kimbrough and Vostruknutov 2016, Barr et al. 2017, Fallucchi and Nosenzo 2020).

on their incentives in the post-task questionnaire, the effects vanish (Figure 2, right panel). Detailed test statistics are reported in Tables A4 and A5 in Appendix A.

Figure 2: Mean appropriateness ratings in *Baseline* and *No Conflict*

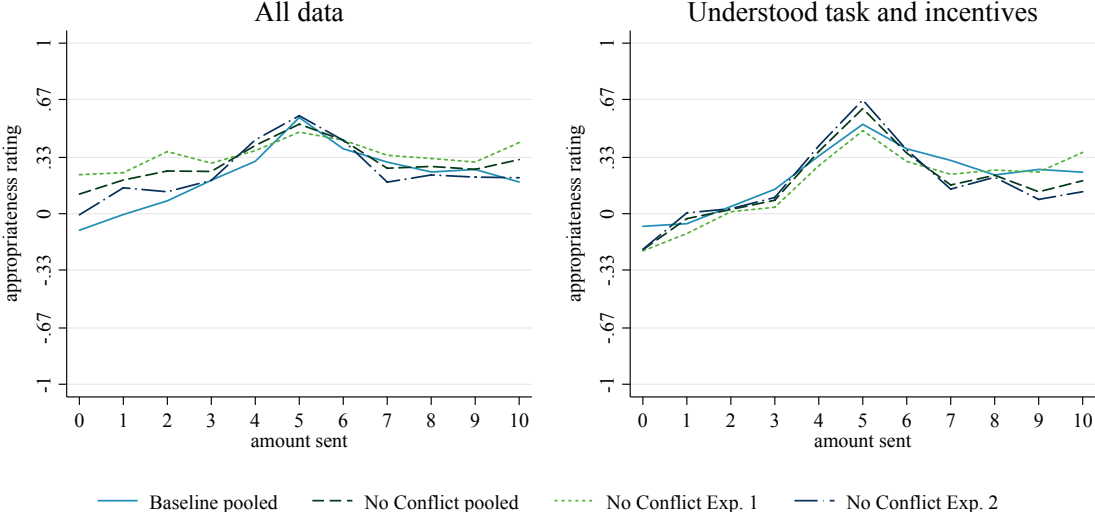
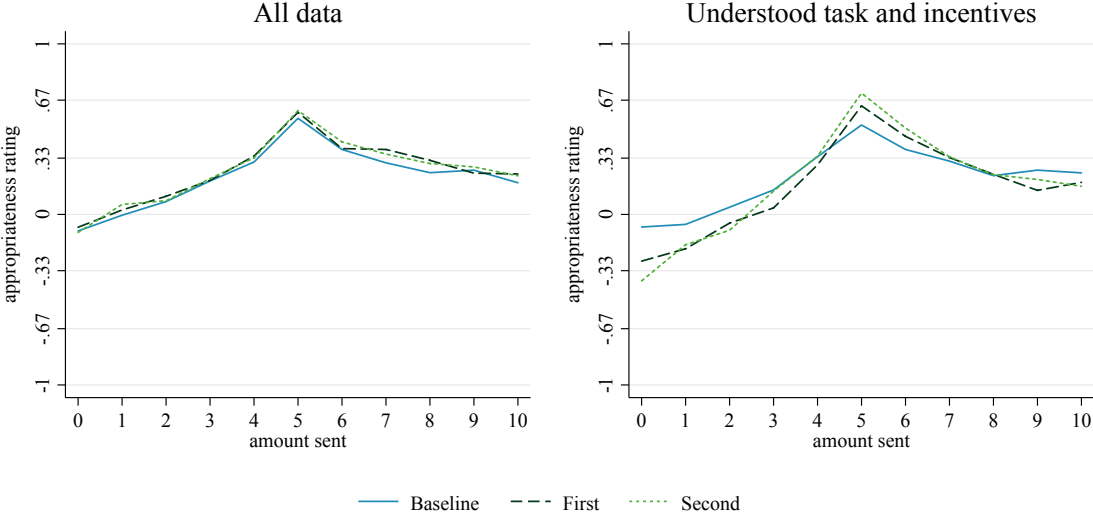


Figure 3: Mean appropriateness ratings in *Baseline*, *First*, and *Second*; pooled data



Removing incentives. In treatments *First* and *Second*, we do not incentivize participants to select the appropriateness rating they think most other participants will select. Instead, we simply ask participants for their first- and second-order beliefs. That is, participants are asked, without incentivization, to rate the appropriateness of the different actions (*First*) and how they think most others would rate them (*Second*), respectively. The responses are depicted in Figure 3. We find that aggregated responses elicited by simply asking for participants’ first- and

second-order beliefs are practically identical to those elicited using KW's incentivized elicitation method. In fact, we do not find any statistically significant differences between these treatments and any of the incentivized treatments. The results hold if we restrict the sample to participants who understood the incentives (or rather lack thereof). Details are reported in Tables A6 and A7 in Appendix A.

4 Discussion and Conclusion

The most-widely used method of eliciting social norms in experiments appears to feature a dissonance between task wording and monetary incentives for the participant. The variety of implementations of KW's task and the various adaptations of the original instructions highlight that it did not go unnoticed. In this study, we attempt to identify the effects this dissonance has on participants' responses. We do not find any effects of reinforcing or attenuating the salience of monetary incentives relative to task wording. Even completely eliminating the dissonance by adapting the instructions does not affect participants' behavior. The apparent dissonance is of no consequence. In addition, we find that response patterns elicited through Krupka and Weber's method are remarkably similar to the patterns of first- and second-order beliefs about the actual appropriateness of the different allocation choices. That is, simply asking participants what they believe most others to think about the appropriateness of an action may yield the same information as implementing the more complicated social norm elicitation method based on coordination. This finding relates to recent work by Fallucchi et al. (2020) and Falk et al. (2018) who demonstrate that behavioral measures can be approximated by survey questions in the context of individuals' willingness to compete and more general economic preferences, respectively.

Conducting the experiment online, we find responses to lie within a relatively small range, resulting in distributions that are flatter than the ones typically found in other experiments implementing the method. We do not think that this is an effect of conducting the experiment online or with the Mechanical Turk sample, as Chang et al. (2019) and Fallucchi and Nosenzo (2020), for example, report larger ranges despite conducting their experiments online with a Mechanical Turk sample. We speculate that the smaller range is the result of presenting the individual allocations separately and in randomized order, rather than in a sorted choice list format.

Finally, we find that proper reading of the instructions does not seem to make a difference with regard to participants' responses. Whether we enforce engagement with the instructions

through pre-task comprehension questions, only look at the sub-sample of participants who correctly answered post-task checks, or simply use all observations does not affect the results. It appears that KW's method is remarkably robust to variations in task wording, incentive salience, and participants' engagement with the tasks' instructions. Yet, implementing KW's method might not always be necessary. Depending on a study's focus, the cheaper, quicker, and easier to explain approach of asking participants for their (non-incentivized) beliefs might be a viable alternative.

Pre-registration and data availability

Pre-registrations, experimental instructions, data, and analysis files are available at https://osf.io/6sakb/?view_only=81ec23ab7234488fb91b8f1c04a6bc30.

Declarations of interest

The author declares not to have any competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Abbink, K., Gangadharan, L., Handfield, T., and Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications*, 8, 609.
- Barr, A., Lane, T., and Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164, 153-164.
- Chang, D., Chen, R., Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 116, 158-178.
- Chen, D.L., Schonger, M., Wickens, C. (2016). oTree – An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9, 88-97.
- D’Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62, 1-7.
- Erkut, H., Nosenzo, D., and Sefton, M. (2015). Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters*, 130, 28-31.
- Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence of Economic Preferences. *The Quarterly Journal of Economics*, 113(4), 1645-1692.
- Fallucchi, F. and Nosenzo, D. (2020). The coordinating power of social norms. Aarhus University Economics Working Papers 2020-06.
- Fallucchi, F., Nosenzo, D., and Reuben, E. (2020). Measuring preferences for competition with experimentally-validated survey questions. *Journal of Economic Behavior & Organization*, 178, 402-423.
- Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms and fair sharing. *European Economic Review*, 97, 72-86.
- Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer Effects in Pro-Social behavior: Social Norms or Social Preferences? *Journal of the European Economic Association*, 11(3), 548-573.
- Huber, C. and Huber, J. (2020). Bad bankers no more? Truth-telling and (dis)honesty in the finance industry. *Journal of Economic Behavior & Organization*, 180, 472-493.
- Kimbrough, E.O. and Vostroknutov, A. (2016). Norms Make Preferences Social. *Journal of the European Economic Association*, 14(3), 608-638.
- Kölle, F., Lane, T., Nosenzo, D., and Starmer, C. (2020). Promoting voter registration: the effects of low-cost interventions on behavior and norms. *Behavioral Public Policy*, 4(1), 26-49.
- Krupka, E.L. and Weber, R.A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495-524.

- Krupka, E.L., Leider, S., and Jiang M. (2017). A Meeting of the minds: Informal Agreements and Social Norms. *Management Science*, 63(6), 1657-2048.
- Sass, M., Timme, F., and Weimann, J. (2018). On the Dynamics of Altruistic Behavior. Working Paper.
- Heinicke, F., König-Kersting, C., and Schmidt, R. (2020). Injunctive vs. Descriptive Social Norms and Reference Group Dependence. Mimeo.
- Vesely, S. (2015). Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision Making*, 10(2), 191-197.
- Vesely, S. and Klöckner, C.A. (2018). Global Social Norms and Environmental Behavior. *Environment and Behavior*, 50(3), 247-272.
- Zhang, P., Zhang, Y., and Palma, M. (2018). Social norms and competitiveness: My willingness to compete depends on who I am (supposed to be). MPRA Working Paper No. 89727.

Appendix A

Table A1: Demographics

Treatment	Observations	Age	Female	Economist
<i>Experiment 1</i>				
Baseline	102	38.28 (11.09)	41.18%	1.96%
Always	106	37.03 (10.86)	43.40%	2.83%
Never	101	37.08 (10.24)	32.67%	1.98%
No Conflict	108	38.55 (10.94)	34.26%	1.85%
First	103	38.47 (11.50)	29.13%	3.88%
Second	100	38.22 (11.01)	29.00%	3.00%
<i>Experiment 2</i>				
Baseline	101	36.68 (9.71)	41.58%	0.00%
Always	104	37.60 (10.94)	42.31%	0.96%
Never	102	36.72 (10.29)	36.27%	3.92%
No Conflict	101	35.19 (10.35)	46.53%	1.98%
First	101	38.90 (10.36)	41.58%	1.98%
Second	99	37.26 (9.49)	37.37%	2.02%

Note: The table shows the number of observations in each treatment, average age of the participants with standard deviation in parenthesis, the share of female participants, and the share of participants who reported to have had some formal training in economics on a university / college level.

Table A2: Appropriateness ratings *Baseline* vs. *Always*, all data

action	<i>Baseline</i>							<i>Always</i>							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>Exp 1</i>	<i>N=102</i>							<i>N=106</i>							
send 0	-0.13	0.25	0.16	0.12	0.23	0.17	0.08	0.00	0.25	0.09	0.09	0.19	0.25	0.11	0.188
send 1	-0.03	0.13	0.21	0.19	0.16	0.25	0.08	0.03	0.11	0.15	0.19	0.25	0.25	0.06	0.620
send 2	0.10	0.01	0.26	0.16	0.21	0.28	0.08	0.23	0.04	0.14	0.13	0.29	0.26	0.13	0.153
send 3	0.17	0.02	0.13	0.26	0.21	0.26	0.12	0.22	0.03	0.08	0.21	0.30	0.30	0.08	0.641
send 4	0.28	0.02	0.05	0.22	0.31	0.31	0.09	0.40	0.02	0.03	0.09	0.39	0.39	0.08	0.106
send 5	0.56	0.01	0.01	0.13	0.22	0.27	0.36	0.51	0.03	0.03	0.08	0.30	0.21	0.35	0.502
send 6	0.38	0.01	0.05	0.16	0.34	0.25	0.19	0.35	0.04	0.07	0.12	0.30	0.30	0.17	0.946
send 7	0.30	0.01	0.10	0.2	0.24	0.30	0.16	0.24	0.02	0.11	0.18	0.34	0.21	0.14	0.315
send 8	0.30	0.04	0.09	0.16	0.25	0.29	0.17	0.22	0.03	0.12	0.21	0.21	0.30	0.13	0.402
send 9	0.30	0.05	0.12	0.11	0.23	0.32	0.18	0.18	0.10	0.08	0.16	0.28	0.21	0.16	0.128
send 10	0.21	0.11	0.13	0.12	0.19	0.24	0.23	0.14	0.16	0.09	0.10	0.23	0.28	0.13	0.303
<i>Exp 2</i>	<i>N=101</i>							<i>N=104</i>							
send 0	-0.06	0.22	0.14	0.13	0.21	0.25	0.06	0.05	0.22	0.05	0.14	0.21	0.28	0.10	0.263
send 1	0.02	0.17	0.16	0.11	0.23	0.24	0.10	0.15	0.12	0.12	0.13	0.26	0.22	0.15	0.195
send 2	0.05	0.05	0.21	0.19	0.25	0.26	0.05	0.20	0.03	0.14	0.18	0.24	0.30	0.11	0.068
send 3	0.22	0.03	0.05	0.25	0.33	0.27	0.08	0.33	.	0.06	0.21	0.28	0.31	0.14	0.105
send 4	0.34	0.01	0.06	0.11	0.43	0.35	0.05	0.38	.	0.03	0.18	0.35	0.30	0.14	0.413
send 5	0.57	0.01	0.02	0.07	0.27	0.33	0.31	0.58	0.01	.	0.10	0.25	0.31	0.34	0.751
send 6	0.39	0.03	0.04	0.10	0.41	0.26	0.17	0.38	0.03	0.04	0.13	0.34	0.30	0.16	0.948
send 7	0.31	0.03	0.10	0.13	0.28	0.34	0.13	0.32	0.04	0.09	0.12	0.33	0.25	0.18	0.881
send 8	0.19	0.06	0.17	0.11	0.28	0.24	0.15	0.31	0.06	0.07	0.14	0.26	0.29	0.18	0.157
send 9	0.21	0.09	0.11	0.13	0.26	0.25	0.17	0.32	0.06	0.12	0.10	0.22	0.28	0.23	0.173
send 10	0.16	0.13	0.13	0.11	0.18	0.32	0.14	0.27	0.13	0.07	0.11	0.19	0.31	0.20	0.246
<i>Pooled</i>	<i>N=203</i>							<i>N=210</i>							
send 0	-0.10	0.24	0.15	0.12	0.22	0.21	0.07	0.02	0.24	0.07	0.12	0.20	0.27	0.10	0.083
send 1	-0.00	0.15	0.18	0.15	0.19	0.24	0.09	0.09	0.11	0.13	0.16	0.25	0.23	0.10	0.209
send 2	0.08	0.03	0.24	0.17	0.23	0.27	0.06	0.21	0.03	0.14	0.16	0.27	0.28	0.12	0.021
send 3	0.20	0.02	0.09	0.26	0.27	0.27	0.10	0.28	0.01	0.07	0.21	0.29	0.30	0.11	0.148
send 4	0.31	0.01	0.05	0.16	0.37	0.33	0.07	0.39	0.01	0.03	0.14	0.37	0.34	0.11	0.079
send 5	0.56	0.01	0.01	0.10	0.24	0.30	0.33	0.55	0.02	0.01	0.09	0.28	0.26	0.34	0.774
send 6	0.38	0.02	0.04	0.13	0.37	0.26	0.18	0.36	0.03	0.05	0.13	0.32	0.30	0.17	0.992
send 7	0.30	0.02	0.10	0.16	0.26	0.32	0.14	0.28	0.03	0.10	0.15	0.33	0.23	0.16	0.529
send 8	0.24	0.05	0.13	0.13	0.27	0.27	0.16	0.27	0.04	0.10	0.18	0.23	0.30	0.16	0.654
send 9	0.26	0.07	0.11	0.12	0.24	0.29	0.17	0.25	0.08	0.10	0.13	0.25	0.24	0.20	0.925
send 10	0.19	0.12	0.13	0.11	0.18	0.28	0.18	0.20	0.14	0.08	0.10	0.21	0.30	0.17	0.955

The table consists of three panels. The first shows data from the first experiment, the second shows data from the second, and the third presents pooled data. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments.

Table A3: Appropriateness ratings *Baseline* vs. *Never*, all data

action	<i>Baseline</i>							<i>Never</i>							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>Exp 1</i>	<i>N=102</i>							<i>N=101</i>							
send 0	-0.13	0.25	0.16	0.12	0.23	0.17	0.08	0.06	0.2	0.12	0.13	0.13	0.26	0.17	0.032*
send 1	-0.03	0.13	0.21	0.19	0.16	0.25	0.08	0.14	0.11	0.14	0.16	0.21	0.19	0.20	0.074
send 2	0.10	0.01	0.26	0.16	0.21	0.28	0.08	0.24	0.04	0.12	0.18	0.24	0.25	0.18	0.073
send 3	0.17	0.02	0.13	0.26	0.21	0.26	0.12	0.28	0.05	0.09	0.18	0.24	0.24	0.21	0.175
send 4	0.28	0.02	0.05	0.22	0.31	0.31	0.09	0.36	0.01	0.04	0.19	0.30	0.32	0.15	0.228
send 5	0.56	0.01	0.01	0.13	0.22	0.27	0.36	0.55	0.01	0.04	0.12	0.17	0.29	0.38	0.898
send 6	0.38	0.01	0.05	0.16	0.34	0.25	0.19	0.42	0.01	0.06	0.15	0.21	0.38	0.20	0.300
send 7	0.30	0.01	0.10	0.20	0.24	0.30	0.16	0.33	0.02	0.09	0.18	0.21	0.33	0.18	0.617
send 8	0.30	0.04	0.09	0.16	0.25	0.29	0.17	0.34	0.02	0.13	0.12	0.27	0.22	0.25	0.578
send 9	0.30	0.05	0.12	0.11	0.23	0.32	0.18	0.29	0.05	0.12	0.12	0.24	0.28	0.20	0.925
send 10	0.21	0.11	0.13	0.12	0.19	0.24	0.23	0.26	0.13	0.07	0.1	0.24	0.24	0.23	0.762
<i>Exp 2</i>	<i>N=101</i>							<i>N=102</i>							
send 0	-0.06	0.22	0.14	0.13	0.21	0.25	0.06	-0.06	0.26	0.09	0.13	0.23	0.18	0.12	0.979
send 1	0.02	0.17	0.16	0.11	0.23	0.24	0.10	0.03	0.16	0.13	0.17	0.23	0.22	0.11	0.932
send 2	0.05	0.05	0.21	0.19	0.25	0.26	0.05	0.09	0.08	0.10	0.23	0.33	0.22	0.05	0.780
send 3	0.22	0.03	0.05	0.25	0.33	0.27	0.08	0.19	0.03	0.08	0.24	0.33	0.25	0.08	0.685
send 4	0.34	0.01	0.06	0.11	0.43	0.35	0.05	0.34	0.02	0.06	0.14	0.37	0.27	0.14	0.750
send 5	0.57	0.01	0.02	0.07	0.27	0.33	0.31	0.56	0.02	0.02	0.06	0.29	0.29	0.31	0.879
send 6	0.39	0.03	0.04	0.10	0.41	0.26	0.17	0.40	0.02	0.05	0.09	0.37	0.33	0.14	0.766
send 7	0.31	0.03	0.10	0.13	0.28	0.34	0.13	0.24	0.03	0.07	0.23	0.30	0.25	0.12	0.299
send 8	0.19	0.06	0.17	0.11	0.28	0.24	0.15	0.22	0.05	0.14	0.14	0.28	0.25	0.15	0.787
send 9	0.21	0.09	0.11	0.13	0.26	0.25	0.17	0.18	0.10	0.16	0.10	0.25	0.20	0.20	0.746
send 10	0.16	0.13	0.13	0.11	0.18	0.32	0.14	0.15	0.17	0.03	0.19	0.21	0.25	0.17	0.914
<i>Pooled</i>	<i>N=203</i>							<i>N=203</i>							
send 0	-0.10	0.24	0.15	0.12	0.22	0.21	0.07	0.00	0.23	0.10	0.13	0.18	0.22	0.14	0.126
send 1	-0.00	0.15	0.18	0.15	0.19	0.24	0.09	0.08	0.13	0.13	0.16	0.22	0.20	0.15	0.187
send 2	0.08	0.03	0.24	0.17	0.23	0.27	0.06	0.16	0.06	0.11	0.20	0.29	0.23	0.11	0.147
send 3	0.20	0.02	0.09	0.26	0.27	0.27	0.10	0.24	0.04	0.08	0.21	0.29	0.24	0.14	0.451
send 4	0.31	0.01	0.05	0.16	0.37	0.33	0.07	0.35	0.01	0.05	0.16	0.33	0.30	0.14	0.270
send 5	0.56	0.01	0.01	0.10	0.24	0.30	0.33	0.55	0.01	0.03	0.09	0.23	0.29	0.34	0.992
send 6	0.38	0.02	0.04	0.13	0.37	0.26	0.18	0.41	0.01	0.05	0.12	0.29	0.35	0.17	0.332
send 7	0.30	0.02	0.1	0.16	0.26	0.32	0.14	0.28	0.02	0.08	0.20	0.26	0.29	0.15	0.734
send 8	0.24	0.05	0.13	0.13	0.27	0.27	0.16	0.28	0.03	0.13	0.13	0.28	0.23	0.20	0.588
send 9	0.26	0.07	0.11	0.12	0.24	0.29	0.17	0.23	0.07	0.14	0.11	0.25	0.24	0.20	0.769
send 10	0.19	0.12	0.13	0.11	0.18	0.28	0.18	0.20	0.15	0.05	0.14	0.22	0.24	0.20	0.884

The table consists of three panels. The first shows data from the first experiment, the second shows data from the second, and the third presents pooled data. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments, * denotes significance at 5%.

Table A4: Appropriateness ratings *Baseline* vs. *No Conflict*, all data

action	<i>Baseline</i>							<i>No Conflict</i>							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>Exp 1</i>	<i>N=102</i>							<i>N=108</i>							
send 0	-0.13	0.25	0.16	0.12	0.23	0.17	0.08	0.23	0.12	0.04	0.13	0.30	0.30	0.12	<0.001***
send 1	-0.03	0.13	0.21	0.19	0.16	0.25	0.08	0.24	0.07	0.09	0.11	0.32	0.28	0.12	0.007**
send 2	0.10	0.01	0.26	0.16	0.21	0.28	0.08	0.36	0.03	0.09	0.11	0.27	0.30	0.20	0.002**
send 3	0.17	0.02	0.13	0.26	0.21	0.26	0.12	0.30	0.01	0.08	0.19	0.31	0.25	0.16	0.163
send 4	0.28	0.02	0.05	0.22	0.31	0.31	0.09	0.37	0.01	0.02	0.14	0.50	0.19	0.15	0.472
send 5	0.56	0.01	0.01	0.13	0.22	0.27	0.36	0.48	0.02	0.03	0.06	0.36	0.33	0.19	0.047
send 6	0.38	0.01	0.05	0.16	0.34	0.25	0.19	0.43	0.01	0.06	0.09	0.37	0.21	0.25	0.417
send 7	0.30	0.01	0.10	0.20	0.24	0.30	0.16	0.34	0.05	0.05	0.13	0.31	0.31	0.15	0.671
send 8	0.30	0.04	0.09	0.16	0.25	0.29	0.17	0.32	0.03	0.08	0.13	0.33	0.26	0.17	0.956
send 9	0.30	0.05	0.12	0.11	0.23	0.32	0.18	0.30	0.06	0.08	0.13	0.24	0.29	0.19	0.993
send 10	0.21	0.11	0.13	0.12	0.19	0.24	0.23	0.42	0.06	0.02	0.08	0.31	0.36	0.17	0.149
<i>Exp 2</i>	<i>N=101</i>							<i>N=101</i>							
send 0	-0.06	0.22	0.14	0.13	0.21	0.25	0.06	-0.01	0.30	0.04	0.10	0.22	0.21	0.14	0.651
send 1	0.02	0.17	0.16	0.11	0.23	0.24	0.10	0.15	0.15	0.12	0.07	0.24	0.31	0.12	0.193
send 2	0.05	0.05	0.21	0.19	0.25	0.26	0.05	0.13	0.09	0.15	0.15	0.26	0.23	0.13	0.361
send 3	0.22	0.03	0.05	0.25	0.33	0.27	0.08	0.20	0.03	0.06	0.29	0.27	0.26	0.10	0.813
send 4	0.34	0.01	0.06	0.11	0.43	0.35	0.05	0.43	.	0.02	0.12	0.34	0.46	0.07	0.072
send 5	0.57	0.01	0.02	0.07	0.27	0.33	0.31	0.57	0.01	0.01	0.08	0.28	0.30	0.33	0.923
send 6	0.39	0.03	0.04	0.10	0.41	0.26	0.17	0.43	.	0.02	0.18	0.22	0.47	0.12	0.318
send 7	0.31	0.03	0.10	0.13	0.28	0.34	0.13	0.19	0.02	0.11	0.25	0.28	0.24	0.11	0.103
send 8	0.19	0.06	0.17	0.11	0.28	0.24	0.15	0.23	0.04	0.13	0.14	0.34	0.21	0.15	0.829
send 9	0.21	0.09	0.11	0.13	0.26	0.25	0.17	0.21	0.11	0.10	0.08	0.30	0.30	0.12	0.884
send 10	0.16	0.13	0.13	0.11	0.18	0.32	0.14	0.21	0.14	0.06	0.08	0.35	0.23	0.15	0.975
<i>Pooled</i>	<i>N=203</i>							<i>N=209</i>							
send 0	-0.10	0.24	0.15	0.12	0.22	0.21	0.07	0.12	0.21	0.04	0.11	0.26	0.25	0.13	0.004**
send 1	-0.00	0.15	0.18	0.15	0.19	0.24	0.09	0.20	0.11	0.11	0.09	0.28	0.29	0.12	0.006**
send 2	0.08	0.03	0.24	0.17	0.23	0.27	0.06	0.25	0.06	0.12	0.13	0.26	0.26	0.17	0.003**
send 3	0.20	0.02	0.09	0.26	0.27	0.27	0.10	0.25	0.02	0.07	0.24	0.29	0.25	0.13	0.369
send 4	0.31	0.01	0.05	0.16	0.37	0.33	0.07	0.40	0.00	0.02	0.13	0.42	0.32	0.11	0.106
send 5	0.56	0.01	0.01	0.10	0.24	0.30	0.33	0.52	0.01	0.02	0.07	0.32	0.32	0.26	0.160
send 6	0.38	0.02	0.04	0.13	0.37	0.26	0.18	0.43	0.00	0.04	0.13	0.30	0.33	0.19	0.212
send 7	0.30	0.02	0.1	0.16	0.26	0.32	0.14	0.27	0.03	0.08	0.19	0.30	0.28	0.13	0.417
send 8	0.24	0.05	0.13	0.13	0.27	0.27	0.16	0.28	0.03	0.11	0.13	0.33	0.23	0.16	0.838
send 9	0.26	0.07	0.11	0.12	0.24	0.29	0.17	0.26	0.09	0.09	0.11	0.27	0.29	0.16	0.908
send 10	0.19	0.12	0.13	0.11	0.18	0.28	0.18	0.32	0.10	0.04	0.08	0.33	0.30	0.16	0.261

The table consists of three panels. The first shows data from the first experiment, the second shows data from the second, and the third presents pooled data. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments, *** / ** denote significance at 1% / 5%.

Table A5: Appropriateness ratings *Baseline* vs. *No Conflict*, understood incentives only

action	<i>Baseline</i>							<i>No Conflict</i>							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>Exp 1</i>	<i>N=54</i>							<i>N=42</i>							
send 0	-0.23	0.33	0.15	0.09	0.19	0.22	0.02	-0.05	0.26	0.05	0.12	0.31	0.24	0.02	0.277
send 1	-0.09	0.22	0.17	0.13	0.17	0.26	0.06	-0.04	0.14	0.19	0.14	0.29	0.17	0.07	0.784
send 2	0.06	.	0.30	0.17	0.20	0.26	0.07	0.12	0.07	0.19	0.12	0.31	0.17	0.14	0.751
send 3	0.17	0.02	0.15	0.22	0.26	0.24	0.11	0.10	0.02	0.14	0.26	0.33	0.14	0.10	0.479
send 4	0.33	.	0.04	0.22	0.28	0.37	0.09	0.31	0.02	0.02	0.17	0.48	0.19	0.12	0.550
send 5	0.61	.	.	0.13	0.19	0.30	0.39	0.48	.	0.02	0.10	0.36	0.33	0.19	0.063
send 6	0.38	.	0.06	0.19	0.28	0.28	0.20	0.33	.	0.12	0.12	0.36	0.21	0.19	0.586
send 7	0.38	.	0.07	0.19	0.24	0.28	0.22	0.25	0.07	0.05	0.17	0.31	0.29	0.12	0.266
send 8	0.31	0.04	0.07	0.15	0.31	0.24	0.19	0.29	0.05	0.07	0.12	0.40	0.21	0.14	0.669
send 9	0.32	0.06	0.11	0.13	0.17	0.31	0.22	0.23	0.10	0.12	0.14	0.10	0.38	0.17	0.600
send 10	0.28	0.06	0.15	0.11	0.19	0.26	0.24	0.36	0.10	0.02	0.10	0.26	0.33	0.19	0.803
<i>Exp 2</i>	<i>N=71</i>							<i>N=72</i>							
send 0	-0.20	0.31	0.15	0.10	0.18	0.21	0.04	-0.15	0.38	0.04	0.10	0.21	0.19	0.08	0.866
send 1	-0.12	0.24	0.18	0.11	0.20	0.17	0.10	0.02	0.21	0.14	0.07	0.22	0.26	0.10	0.265
send 2	-0.06	0.06	0.30	0.18	0.21	0.21	0.04	0.05	0.10	0.19	0.14	0.26	0.21	0.10	0.333
send 3	0.15	0.04	0.07	0.28	0.31	0.21	0.08	0.12	0.04	0.07	0.33	0.25	0.21	0.10	0.830
send 4	0.33	0.01	0.06	0.13	0.38	0.38	0.04	0.40	.	0.01	0.14	0.42	0.35	0.08	0.530
send 5	0.59	0.01	0.03	0.07	0.23	0.27	0.39	0.67	.	.	0.07	0.22	0.28	0.43	0.483
send 6	0.35	0.03	0.06	0.13	0.39	0.20	0.20	0.38	.	0.01	0.21	0.26	0.44	0.07	0.748
send 7	0.21	0.04	0.14	0.17	0.23	0.30	0.13	0.17	0.03	0.11	0.25	0.29	0.19	0.13	0.578
send 8	0.14	0.08	0.20	0.10	0.25	0.18	0.18	0.22	0.06	0.14	0.13	0.29	0.24	0.15	0.568
send 9	0.16	0.11	0.14	0.13	0.21	0.20	0.21	0.10	0.15	0.14	0.06	0.32	0.24	0.10	0.363
send 10	0.08	0.17	0.15	0.08	0.18	0.25	0.15	0.14	0.18	0.08	0.07	0.31	0.18	0.18	0.829
<i>Pooled</i>	<i>N=125</i>							<i>N=114</i>							
send 0	-0.22	0.32	0.15	0.10	0.18	0.22	0.03	-0.11	0.33	0.04	0.11	0.25	0.21	0.06	0.425
send 1	-0.11	0.23	0.18	0.12	0.18	0.21	0.08	-0.00	0.18	0.16	0.10	0.25	0.23	0.09	0.281
send 2	-0.01	0.03	0.30	0.18	0.21	0.23	0.06	0.08	0.09	0.19	0.13	0.28	0.19	0.11	0.380
send 3	0.15	0.03	0.10	0.26	0.29	0.22	0.10	0.11	0.04	0.10	0.31	0.28	0.18	0.10	0.544
send 4	0.33	0.01	0.05	0.17	0.34	0.38	0.06	0.37	0.01	0.02	0.15	0.44	0.29	0.10	0.925
send 5	0.60	0.01	0.02	0.10	0.21	0.28	0.39	0.60	.	0.01	0.08	0.27	0.30	0.34	0.602
send 6	0.37	0.02	0.06	0.15	0.34	0.23	0.20	0.36	.	0.05	0.18	0.30	0.36	0.11	0.890
send 7	0.28	0.02	0.11	0.18	0.23	0.29	0.17	0.20	0.04	0.09	0.22	0.30	0.23	0.12	0.206
send 8	0.22	0.06	0.14	0.12	0.28	0.21	0.18	0.24	0.05	0.11	0.12	0.33	0.23	0.15	0.915
send 9	0.23	0.09	0.13	0.13	0.19	0.25	0.22	0.15	0.13	0.13	0.09	0.24	0.29	0.12	0.222
send 10	0.17	0.12	0.15	0.10	0.18	0.26	0.19	0.22	0.15	0.06	0.08	0.29	0.24	0.18	0.849

The table consists of three panels. The first shows data from the first experiment, the second shows data from the second, and the third presents pooled data. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments.

Table A6: Appropriateness ratings *Baseline* vs. *First*

action	Baseline							First							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>All data</i>	<i>N=203</i>							<i>N=204</i>							
send 0	-0.10	0.24	0.15	0.12	0.22	0.21	0.07	-0.08	0.27	0.10	0.11	0.22	0.18	0.11	0.819
send 1	-0.00	0.15	0.18	0.15	0.19	0.24	0.09	0.03	0.13	0.17	0.15	0.24	0.20	0.11	0.723
send 2	0.08	0.03	0.24	0.17	0.23	0.27	0.06	0.11	0.04	0.18	0.21	0.25	0.23	0.10	0.581
send 3	0.20	0.02	0.09	0.26	0.27	0.27	0.10	0.20	0.02	0.11	0.22	0.28	0.27	0.09	0.996
send 4	0.31	0.01	0.05	0.16	0.37	0.33	0.07	0.34	0.01	0.06	0.17	0.32	0.31	0.13	0.333
send 5	0.56	0.01	0.01	0.10	0.24	0.30	0.33	0.60	0.01	0.02	0.06	0.25	0.3	0.35	0.531
send 6	0.38	0.02	0.04	0.13	0.37	0.26	0.18	0.39	0.02	0.05	0.13	0.29	0.36	0.15	0.640
send 7	0.30	0.02	0.1	0.16	0.26	0.32	0.14	0.38	0.00	0.07	0.14	0.32	0.27	0.19	0.255
send 8	0.24	0.05	0.13	0.13	0.27	0.27	0.16	0.32	0.03	0.11	0.13	0.25	0.30	0.18	0.216
send 9	0.26	0.07	0.11	0.12	0.24	0.29	0.17	0.24	0.08	0.10	0.14	0.24	0.24	0.20	0.891
send 10	0.19	0.12	0.13	0.11	0.18	0.28	0.18	0.24	0.12	0.07	0.13	0.24	0.24	0.21	0.569
<i>Understood incentives</i>	<i>N=125</i>							<i>N=120</i>							
send 0	-0.22	0.32	0.15	0.10	0.18	0.22	0.03	-0.21	0.38	0.09	0.12	0.16	0.14	0.12	0.958
send 1	-0.11	0.23	0.18	0.12	0.18	0.21	0.08	-0.15	0.18	0.24	0.17	0.17	0.13	0.10	0.781
send 2	-0.01	0.03	0.30	0.18	0.21	0.23	0.06	-0.03	0.07	0.22	0.25	0.22	0.17	0.08	0.886
send 3	0.15	0.03	0.10	0.26	0.29	0.22	0.10	0.11	0.02	0.15	0.26	0.30	0.19	0.08	0.453
send 4	0.33	0.01	0.05	0.17	0.34	0.38	0.06	0.32	0.02	0.06	0.15	0.37	0.31	0.10	0.867
send 5	0.60	0.01	0.02	0.10	0.21	0.28	0.39	0.63	0.02	0.01	0.05	0.24	0.28	0.41	0.671
send 6	0.37	0.02	0.06	0.15	0.34	0.23	0.20	0.43	0.03	0.03	0.12	0.29	0.34	0.19	0.256
send 7	0.28	0.02	0.11	0.18	0.23	0.29	0.17	0.39	0.01	0.07	0.13	0.30	0.28	0.20	0.238
send 8	0.22	0.06	0.14	0.12	0.28	0.21	0.18	0.28	0.04	0.13	0.13	0.22	0.29	0.18	0.396
send 9	0.23	0.09	0.13	0.13	0.19	0.25	0.22	0.22	0.09	0.11	0.15	0.23	0.20	0.23	0.873
send 10	0.17	0.12	0.15	0.10	0.18	0.26	0.19	0.23	0.14	0.07	0.12	0.23	0.20	0.25	0.501

The table consists of two panels. The first shows all data and the second shows data from participants who correctly answered the question on understanding the incentives. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments.

Table A7: Appropriateness ratings *Baseline* vs. *Second*

action	Baseline							Second							p-value
	mean	---	--	-	+	++	+++	mean	---	--	-	+	++	+++	
<i>All data</i>	<i>N=203</i>							<i>N=199</i>							
send 0	-0.10	0.24	0.15	0.12	0.22	0.21	0.07	-0.11	0.28	0.12	0.12	0.21	0.19	0.10	0.866
send 1	-0.00	0.15	0.18	0.15	0.19	0.24	0.09	0.06	0.13	0.17	0.13	0.25	0.19	0.13	0.419
send 2	0.08	0.03	0.24	0.17	0.23	0.27	0.06	0.08	0.07	0.15	0.21	0.27	0.23	0.08	0.964
send 3	0.20	0.02	0.09	0.26	0.27	0.27	0.10	0.21	0.02	0.13	0.18	0.29	0.31	0.07	0.873
send 4	0.31	0.01	0.05	0.16	0.37	0.33	0.07	0.33	0.01	0.05	0.16	0.40	0.31	0.09	0.768
send 5	0.56	0.01	0.01	0.10	0.24	0.30	0.33	0.61	0.01	0.01	0.09	0.23	0.30	0.37	0.358
send 6	0.38	0.02	0.04	0.13	0.37	0.26	0.18	0.43	0.01	0.06	0.09	0.35	0.34	0.16	0.346
send 7	0.30	0.02	0.1	0.16	0.26	0.32	0.14	0.35	0.01	0.08	0.15	0.29	0.29	0.18	0.445
send 8	0.24	0.05	0.13	0.13	0.27	0.27	0.16	0.30	0.04	0.11	0.15	0.24	0.27	0.20	0.313
send 9	0.26	0.07	0.11	0.12	0.24	0.29	0.17	0.28	0.06	0.09	0.18	0.22	0.24	0.22	0.633
send 10	0.19	0.12	0.13	0.11	0.18	0.28	0.18	0.23	0.14	0.08	0.12	0.21	0.26	0.21	0.579
<i>Understood incentives</i>	<i>N=125</i>							<i>N=109</i>							
send 0	-0.22	0.32	0.15	0.10	0.18	0.22	0.03	-0.29	0.37	0.15	0.12	0.16	0.16	0.06	0.417
send 1	-0.11	0.23	0.18	0.12	0.18	0.21	0.08	-0.08	0.16	0.20	0.16	0.27	0.13	0.09	0.717
send 2	-0.01	0.03	0.30	0.18	0.21	0.23	0.06	-0.00	0.07	0.19	0.22	0.26	0.21	0.05	0.937
send 3	0.15	0.03	0.10	0.26	0.29	0.22	0.10	0.13	0.02	0.14	0.22	0.34	0.24	0.05	0.674
send 4	0.33	0.01	0.05	0.17	0.34	0.38	0.06	0.35	.	0.06	0.15	0.38	0.33	0.09	0.876
send 5	0.60	0.01	0.02	0.10	0.21	0.28	0.39	0.67	.	.	0.08	0.20	0.27	0.45	0.301
send 6	0.37	0.02	0.06	0.15	0.34	0.23	0.20	0.48	.	0.02	0.10	0.35	0.38	0.16	0.158
send 7	0.28	0.02	0.11	0.18	0.23	0.29	0.17	0.37	0.01	0.08	0.15	0.29	0.25	0.22	0.312
send 8	0.22	0.06	0.14	0.12	0.28	0.21	0.18	0.24	0.05	0.15	0.17	0.18	0.24	0.22	0.566
send 9	0.23	0.09	0.13	0.13	0.19	0.25	0.22	0.24	0.06	0.13	0.17	0.20	0.17	0.27	0.793
send 10	0.17	0.12	0.15	0.10	0.18	0.26	0.19	0.17	0.19	0.07	0.10	0.17	0.20	0.26	0.832

The table consists of two panels. The first shows all data and the second shows data from participants who correctly answered the question on understanding the incentives. In each panel, the shares of participants that gave the respective rating for each amount sent by the dictator is reported. The modal response is highlighted. In addition, the mean appropriateness rating on a scale from -1 to 1 is given. The last column presents p-values of Mann-Whitney-U tests between the two treatments.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

<https://www.uibk.ac.at/eeecon/wopec/>

- 2021-02 **Christian König-Kersting:** On the Robustness of Social Norm Elicitation
- 2021-01 **Laura Hueber, Rene Schwaiger:** Debiasing Through Experience Sampling: The Case of Myopic Loss Aversion.
- 2020-34 **Kai A. Konrad, Florian Morath:** The Volunteer's Dilemma in Finite Populations
- 2020-33 **Katharina Momsen, Markus Ohndorf:** Expressive Voting vs. Self-Serving Ignorance
- 2020-32 **Silvia Angerer, Daniela Glätzle-Rützler, Christian Waibel:** Monitoring institutions in health care markets: Experimental evidence
- 2020-31 **Jana Friedrichsen, Katharina Momsen, Stefano Piasenti:** Ignorance, Intention and Stochastic Outcomes
- 2020-30 **Esther Blanco, Alexandra Baier, Felix Holzmeister, Tarek Jaber-Lopez, Natalie Struwe:** Substitution of social concerns under the Covid-19 pandemic
- 2020-29 **Andreas Hackethal, Michael Kirchler, Christine Laudenbach, Michael Razen, Annika Weber:** On the (ir)relevance of monetary incentives in risk preference elicitation experiments
- 2020-28 **Andrej Gill, Matthias Heinz, Heiner Schumacher, Matthias Sutter:** Trustworthiness in the Financial Industry
- 2020-27 **Matthias Sutter, Michael Weyland, Anna Untertrifaller, Manuel Froitzheim:** Financial literacy, risk and time preferences - Results from a randomized educational intervention
- 2020-26 **Rene Schwaiger, Jürgen Huber, Michael Kirchler, Daniel Kleinlercher, Utz Weitzel:** Unequal Opportunities, Social Groups, and Redistribution
- 2020-25 **Roman Inderst, Martin Obradovits:** Competitive Strategies when Consumers are Relative Thinkers: Implications for Pricing, Promotions, and Product Choice
- 2020-24 **Martin Obradovits, Philipp Plaickner:** Price-Directed Search and Collusion
- 2020-23 **Helena Fornwagner, Oliver P. Hauser:** Climate action for (my) children
- 2020-22 **Esther Blanco, Natalie Struwe, James M. Walker:** Incentivizing public good provision through outsider transfers: experimental evidence on sharing rules and additional requirements

- 2020-21 **Loukas Balafoutas, Helena Fornwagner, Rudolf Kerschbamer, Matthias Sutter, Maryna Tverdostup:** Diagnostic Uncertainty and Insurance Coverage in Credence Goods Markets
- 2020-20 **Anna Ulrichshofer, Markus Walzl:** Customer Disputes, Misconduct, and Reputation Building in the Market for Financial Advice
- 2020-19 **Anna Ulrichshofer, Markus Walzl:** Social Comparison and Optimal Contracts in the Competition for Managerial Talent
- 2020-18 **Martin Obradovits, Philipp Plaickner:** Searching for Treatment
- 2020-17 **Jun Honda:** The Gender-Punishment Gap revisited
- 2020-16 **Jun Honda:** The Relation between Rankings and Risk-Taking in the Labor Market for Financial Advice
- 2020-15 **Christina Bannier, Eberhard Feess, Natalie Packham, Markus Walzl:** Differentiation and Risk-Aversion in Imperfectly Competitive Labor Markets
- 2020-14 **Felix Holzmeister, Rudolf Kerschbamer:** oTree: The Equality Equivalence Test
- 2020-13 **Parampreet Christopher Bindra, Graeme Pearce:** The effect of priming on fraud: Evidence from a natural field experiment
- 2020-12 **Alessandro De Chiara, Marco A. Schwarz:** A Dynamic Theory of Regulatory Capture
- 2020-11 **Christoph Huber, Jürgen Huber, Michael Kirchler:** Market shocks and professionals' investment behavior - Evidence from the COVID-19 crash
- 2020-10 **Elisabeth Gsottbauer, Daniel Müller, Samuel Müller, Stefan T. Trautmann, Galina Zudenkova:** Social class and (un)ethical behavior: Causal versus correlational evidence
- 2020-09 **Parampreet Christopher Bindra, Rudolf Kerschbamer, Daniel Neururer, Matthias Sutter:** Reveal it or conceal it: On the value of second opinions in a low-entry-barriers credence goods market
- 2020-08 **Robert Steiger, Eva Posch, Gottfried Tappeiner, Janette Walde:** Effects of climate change on tourism demand considering individual seasonal preferences
- 2020-07 **Fang Liu, Alexander Rasch, Marco A. Schwarz, Christian Waibel:** The role of diagnostic ability in markets for expert services
- 2020-06 **Matthias Stefan, Jürgen Huber, Michael Kirchler, Matthias Sutter, Markus Walzl:** Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect
- 2020-05 **Michael Razen, Jürgen Huber, Laura Hueber, Michael Kirchler, Matthias Stefan:** Financial Literacy, Economic Preferences, and Adolescents' Field Behavior

- 2020-04 **Christian König-Kersting, Johannes Lohse, Anna Louisa Merkel:** Active and Passive Risk-Taking
- 2020-03 **Christoph Huber, Jürgen Huber:** Bad bankers no more? Truth-telling and (dis)honesty in the finance industry
- 2020-02 **Dietmar Fehr, Daniel Müller, Marcel Preuss:** Social Mobility Perceptions and Inequality Acceptance
- 2020-01 **Loukas Balafoutas, Rudolf Kerschbamer:** Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions

University of Innsbruck

Working Papers in Economics and Statistics

2021-02

Christian König-Kersting

On the Robustness of Social Norm Elicitation

Abstract

We study the robustness of Krupka and Weber's method (2013) for eliciting social norms. In two experiments with more than 1200 participants, we find that participants' response patterns are invariant to differences in the salience of the monetarily incentivized coordination aspect. We further demonstrate that asking participants for their personal first and second order beliefs without monetary incentives results in qualitatively identical responses. In addition, we observe that participants give sensible responses whether or not they understand the task or their monetary incentives. Overall, Krupka and Weber's method produces remarkably robust response patterns.

ISSN 1993-4378 (Print)

ISSN 1993-6885 (Online)