

Measuring social norm variation across contexts: Replication and comparison to alternative methods

David Huffman, Garrett Kohno, Pauline Madiès, Spencer Vogrinec,
Stephanie W. Wang, Dhvani Yagnaraman*

September 2024

Abstract

Studying social norms and how they vary across contexts requires reliable measurements. We revisit the seminal Krupka and Weber (KW 2013) norm elicitation method. Our study assesses the importance of two dimensions of eliciting norms — whether to use the KW coordination game approach or a simple two-stage approach that directly elicits first-order or second-order beliefs about social appropriateness separately, and whether to use financial incentives. We replicate KW’s main finding of a qualitative difference in norms between the dictator game and a re-framed version that involves potentially taking money. All methods show that taking money is less socially appropriate than giving money, holding outcomes fixed and regardless of monetary incentives. However, we find that the difference in elicited social appropriateness between the two contexts varies across methods, with elicited first-order beliefs exhibiting the largest gap in social appropriateness and KW eliciting the smallest gap. A comprehension check reveals that nearly half of KW participants initially misunderstood the task. Our results suggest that complex norm elicitation methods may be less sensitive to the norm-relevant context than measuring first-order or second-order beliefs directly due to attenuation arising largely from strategic uncertainty rather than confusion.

JEL Classifications: C91, D64, D91

Keywords: Social Norms; Norm Elicitation; Incentives; Higher-Order Beliefs

1 Introduction

Injunctive norms, defined as shared perceptions among members of a population regarding the appropriateness of different behaviors, are increasingly viewed as important drivers of economic behavior. They have been argued to be important for prosocial behaviors (Krupka & Weber, 2013; Bicchieri et al., 2022), honesty (Abeleler et al., 2019; Bicchieri et al., 2023), discrimination (Barr et al., 2018), and female labor force participation (Bursztyn et al., 2020), to name a few.

Studying social norms requires reliable measurements, and Krupka and Weber (2013) made a seminal contribution in this regard. Their approach to norm measurement (henceforth the KW method) asks participants about the social appropriateness of different actions in experimental games, and incentivizes them to guess the modal response of other participants. The core result of their study was that participants indicated different levels of social appropriateness of the actions in two games, the classic dictator game, and a re-framed version called the “bully” game. This type of framing had been shown previously to strongly influence behavior (List, 2007). The difference in social norms captured by the KW method offered an explanation and demonstrated how measures of norms could be used to predict behavior.

One goal of our paper is to provide a replication test of the main result of Krupka and Weber (2013). We implement the same two games, using the same parameters and instructions, with a notable difference being that we used a representative, online sample rather than college students.¹ This provides, to our knowledge, the first attempt to replicate this aspect of the original study.

Another goal of our paper is to evaluate the KW method relative to other methods of detecting norm differences across games. One reason to test alternative methods is the conceptual link between the KW method and commonly agreed upon definitions of what a norm is are not entirely clear. In particular,

*We thank Ben Greiner for the inspiration. We thank Luca Braghieri, Erin Krupka, Roberto Weber, and conference and seminar participants at the External Validity, Generalizability and Replicability of Economics Experiment Workshop of the 2024 Barcelona Summer Forum, NHH-Rady Spring School in Behavioral Economics, and the University of Pittsburgh Experimental Economics Brown Bag for their helpful comments.

Huffman: Department of Economics, University of Pittsburgh, Pittsburgh, PA, huffmand@pitt.edu; Kohno: Department of Economics, University of Pittsburgh, Pittsburgh, PA gkh8@pitt.edu; Madiès: Department of Economics, Sciences Po, Paris, France, pauline.madies@sciencespo.fr; Vogrinec: Katz Graduate School of Business, University of Pittsburgh, PA, spencer.vogrinec@pitt.edu; Wang: Department of Economics, University of Pittsburgh, PA, swwang@pitt.edu; Yagnaraman: Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, dyagnara@andrew.cmu.edu

¹We also included a control question between the instructions and the experiment choices, which tests awareness of the rules for incentivization.

injunctive norms are usually conceptualized as second-order beliefs about the first-order beliefs that the typical person has about social appropriateness (Görges and Nosenzo, 2020). One challenge is that the KW method incentivizes participants to guess what others guess, but since everyone faces the same incentive, higher order beliefs potentially come into play (participants need to guess what other participants guess about what other participants guess, and so on). This aspect of the KW method may introduce factors such as strategic uncertainty, which could act to attenuate the measurement of how norms vary across games. Simpler, two-stage versions of norm measures have been proposed, in which a first stage elicits first-order beliefs, and a second stage elicits guesses of these first-order beliefs. Such methods might be more sensitive to norm differences across games by eliminating strategic uncertainty or general confusion. Another important question is how crucial incentives are for delivering reliable measures of norm differences across games. We provide a first comparison of KW versus two-stage methods, and of incentivized versus non-incentivized versions of these, in terms of how they assess norms in the dictator versus bully games.

The first stage of our analysis shows that the core result of Krupka and Weber (2013) replicates. Across the dictator and bully games, there is a highly significant difference in responses using the KW method. Participants indicated higher levels of social appropriateness for not giving in the dictator game, compared to taking in the bully game. This difference, but also the magnitudes of the appropriateness ratings in each game, are quite similar to those found in the original study.

The second stage of our analysis compares how participants respond to different measures of social norms across the dictator and bully games. We find that all measures deliver the same qualitative result on the norm difference across games, regardless of monetary incentives. We find, however, that the magnitude of differences in social appropriateness between the two choice environments varies across methods; the participants who stated their social appropriateness ratings directly in the first stage (the Evaluators) exhibit the largest gap in social appropriateness, and the KW methods exhibit the smallest gap. One potential explanation is that the greater strategic uncertainty inherent in the KW method (having to guess what others guess about what others guess) leads to attenuation in responding to the change in games. Another source of attenuation in the KW method could be confusion. A comprehension check prior to the dictator game suggests that participants have difficulty understanding the KW method due to the dissonance between what the original instructions ask participants to do and what is incentivized.

Our paper contributes to a previous literature studying the KW method and alternative methods of measuring norms. Bichieri and Xiao (2009) use a two-stage method to study how beliefs about what others think is appropriate in the dictator game, and beliefs about how others behave in the dictator game, influence one’s own behavior in the dictator game. Fallucchi and Nosenzo (2020) raise the possibility that salient focal points could skew answers in the KW method, because the incentives implement a pure coordination game with multiple equilibria. Focusing on the dictator game, they show that introducing focal visual labels for some of the appropriateness levels does influence responses in the KW measure. Koenig-Kersting (2024) tests whether modifications to the KW instructions designed to make incentives more or less salient and to reduce confusion, affect responses in the dictator game, and also varies whether there are real incentives. The study also compares the KW method to a two-stage method. In that study, none of these variations makes any differences for responses about the dictator game. Our study is complementary to this literature because it offers a replication, and comparison, of how alternative methods of norm measurement do in terms of capturing norm differences across games. While we find little difference across measurement approaches for the dictator game, similar to some previous findings, comparing to the bully game reveals a potentially important method variance, consistent with increasing attenuation as measures become progressively more complex and strategically uncertain.

This paper is organized as follows. Section 2 describes our online experiment that tests five elicitation methods. Section 3 presents results. Section 4 concludes.

2 Experiment

2.1 Design

Table 1 summarizes the treatment conditions, which vary the type of beliefs elicited and the presence of monetary incentives.² Participants in the *Evaluators* condition are instructed to submit first-order beliefs about the social appropriateness of taking certain actions in the dictator game. Participants in the *Second-Stage (Non-Incentivized Second-Stage)* condition are instructed to guess the actual (non-incentivized) modal belief of the *Evaluators* by forming second-order beliefs. Participants in the *Krupka-Weber* condition are given the same instructions as the *Evaluators*, except they are incentivized to guess the modal belief of others

²The instructions for each condition’s task can be found in Table 4 in the appendix.

in the study, thereby forming higher-order beliefs. Finally, participants in the *Non-Incentivized Krupka-Weber* condition are instructed to guess the modal belief of others in the study, again forming higher-order beliefs.

Table 1: Experimental Conditions

Treatment	Elicited Beliefs	Monetary Incentive
<i>Evaluators</i>	First-order	No
<i>Second-Stage</i>	Second-order	Yes
<i>Non-Incentivized Second-Stage</i>	Second-order	No
<i>Krupka-Weber</i>	Higher-order	Yes
<i>Non-Incentivized Krupka-Weber</i>	Higher-order	No

Within each elicitation method, participants are randomly assigned to receive either the “standard” or the “bully” choice environment of the dictator game from Krupka & Weber (2013). In the standard game, participants are presented with a scenario where a dictator is endowed with \$10 and can choose to give any amount of this money, in one-dollar increments to the recipient, who initially receives \$0. An example of an action might be for Individual A (dictator) to “Give \$4 to Individual B. Individual A gets \$6, Individual B gets \$4,” which is then represented as (\$6, \$4). In the bully game, participants are presented with a scenario where both the dictator and the recipient initially receive \$5. In this choice environment, the dictator has the opportunity to not only *give*, but also to *take* money from the recipient in one-dollar increments.

The two alternative choice environments are such that the set of potential payoffs are the same, while the contextual features of the action sets differ. To illustrate this, the corresponding action in the bully game that leads to the same outcome in the previous example from the standard game is for Individual A (dictator) to “Take \$1 from Individual B. Individual A gets \$6, Individual B gets \$4,” which is then represented as (\$6,\$4). Although both actions yield the same payoffs, they are governed by different social norms, with actions involving taking being considered generally less socially appropriate than actions involving giving (Krupka & Weber 2013).

2.2 Procedure

The experiment was conducted on the survey platform Prolific using a representative U.S. sample balanced on age, gender, and political affiliation. The sample comprised of 751 participants. Data collection took place in January, 2024. The experiment was coded in oTree (Chen et al. 2016).

The experiment begins with a set of instructions that vary by elicitation method (see Appendix). To ensure that participants understand the instructions, they are provided with an example situation. After participants respond to the example situation, they receive an explanation of how they could have responded based on their elicitation method.

The example is followed by a comprehension question, which simply asks how participants should indicate their responses. The possible answer options displayed for this question differ based on the assigned elicitation method. All participants see two options: (i) reporting first order beliefs and (ii) an obviously incorrect option to ensure they are paying attention. Those assigned to the *Second-Stage* or *Krupka-Weber* elicitation methods (including non-incentivized versions) also see a third option about reporting second order or higher order beliefs respectively. This differentiation ensures that participants do not become aware of different elicitation methods and do not get confused between the method they are assigned to and another method. Participants answer the comprehension question as many times as needed to get the correct answer.

Participants are then presented with the main norm elicitation task for either the standard or bully variant of the dictator game. In one of the choice environments, participants rate the social appropriateness of each possible action on a four-point scale ranging from “Very socially inappropriate” to “Very socially appropriate”,³ in line with their assigned elicitation method. After participants indicated social appropriateness ratings for all of the eleven possible dictator actions, the experimenter randomly selected one possible action in each variant of the game. Participants in the *Second-Stage* and *Krupka-Weber* conditions received a bonus payment of \$2, in addition to a \$3 participation payment that all participants received, if they had selected the modal appropriateness rating in the *Evaluators* and *Krupka-Weber* conditions for the selected action.⁴

³The scale comprises four possible ratings: “Very socially inappropriate” (coded as -1), “Somewhat socially inappropriate” (coded as -0.3), “Somewhat socially appropriate” (coded as 0.3) and “Very socially appropriate” (coded as 1).

⁴Across all treatments, the average earnings were \$3.73, while the average earnings per hour were \$16.54.

3 Results

3.1 Identifying taking vs. giving norms

First, we investigate whether the two dictator game variants are governed by a different set of norms. We expect actions that leave the dictator with more money than the recipient, namely outcomes (\$10,\$0) to (\$6,\$4), to be associated with lower social appropriateness ratings in the bully game than in the standard game. These actions involve *taking* money from the recipient in the bully game, which is considered to be less socially appropriate than actions in the standard game that involve *giving* money to the recipient.

Table 2 compares social appropriateness ratings elicited by the different methods across the two choice environments. The first column reports differences in social appropriateness ratings across games and methods for outcomes (\$10, \$0) to (\$6, \$4). We replicate Krupka & Weber’s (2013) findings of a difference in norms between “standard” and “bully” dictator games. Robust to all elicitation methods, taking money from is considered less socially appropriate than giving money to the recipient when the two choices result in the same monetary outcome. Indeed, social appropriateness of the taking environment is significantly less than that of the giving environment in outcomes (\$10, \$0) to (\$6, \$4) and not statistically different for outcomes (\$5, \$5) to (\$0, \$10).

Table 2: Difference in Giving and Taking Norms

	Allocation Outcomes	
	(\$10,\$0) - (\$6,\$4)	(\$5,\$5) - (\$0,\$10)
Bully	-0.221*** (0.059)	0.066 (0.087)
Bully × Evaluators	-0.205** (0.096)	-0.149 (0.126)
Bully × NI Second-Stage	-0.049 (0.085)	0.091 (0.122)
Bully × Second-Stage	-0.140 (0.094)	0.000 (0.123)
Bully × NI Krupka-Weber	0.055 (0.088)	-0.118 (0.126)
Evaluators	0.168** (0.074)	0.110 (0.091)
NI Second-Stage	0.014 (0.059)	-0.096 (0.086)
Second-Stage	0.141* (0.072)	0.055 (0.088)
NI Krupka-Weber	0.020 (0.063)	0.011 (0.090)
Constant	-0.420*** (0.041)	0.297*** (0.064)
Observations	3,755	4,506
R^2	0.07	0.01

The reference group for the allocation game is the “Standard” variant. The reference group for the elicitation method is *Krupka-Weber*. Standard errors are clustered at the individual level.

3.2 Incentives

Table 3 reports the effect of monetary incentives on social appropriateness ratings. Across both games and all outcomes, the coefficient on incentives is not different from 0, suggesting that monetary incentives do not yield any significant differences in elicited norms between methods. Additionally, no systematically significant differences between the variances of ratings exist between the incentivized and non-incentivized methods.⁵ Following these results, incentivized and non-incentivized methods are pooled together from this point onwards.

⁵Table 5 in the Appendix presents the results of F-tests comparing the variance in ratings of incentivized and non-incentivized participants for both games and all outcomes.

Table 3: Presence of incentives on dictator game norms

	Standard		Bully	
	(\$10,\$0) - (\$6,\$4)	(\$5,\$5) - (\$0,\$10)	(\$10,\$0) - (\$6,\$4)	(\$5,\$5) - (\$0,\$10)
Two Stage	-0.006 (0.064)	-0.107 (0.085)	-0.110* (0.062)	0.101 (0.092)
Incentivized	-0.020 (0.063)	-0.011 (0.090)	-0.074 (0.061)	0.107 (0.088)
Two Stage \times Incentivized	0.146 (0.096)	0.162 (0.122)	0.110 (0.086)	-0.047 (0.126)
Constant	-0.400*** (0.047)	0.308*** (0.063)	-0.566*** (0.044)	0.256*** (0.066)
Observations	1,595	1,914	1,405	1,686
R^2	0.01	0.01	0.01	0.01

The reference group for the elicitation method is *Non-Incentivized Krupka-Weber*. Standard errors are clustered at the individual level.

3.3 Gap in Social Appropriateness

Our results suggests that all methods identify a difference in social norms for outcomes (\$10, \$0) to (\$6, \$4), where taking is uniformly considered less socially appropriate than giving. However, a difference in the magnitude of this gap in social appropriateness emerges between elicitation methods. Figure 1 shows social appropriateness ratings for the two games by elicitation method. The difference in norm ratings across games is the largest when we elicit first-order beliefs, namely the judgments about social appropriateness of the evaluators in the two stage method, and the smallest when we elicit higher-order beliefs using the KW method. The difference in gaps is statistically significant for outcomes (\$10, \$0) to (\$6, \$4), as reported in Table 2.

Attenuation of the taking-giving gap under the KW method could be due to judging each action on the resulting monetary payoffs, rather than on other contextual features of the games. This might arise from strategic uncertainty or confusion. There is inherently strategic uncertainty in the KW methods as participants are guessing each others' guesses about each others' guesses, rather than simply guessing what others think is socially appropriate. The uncertainty present in others' responses could lead to attenuation through the deliberate use of monetary payoffs as a coordination device. Furthermore, due to the dissonance between the task instructions and incentives under the KW method, there might also be confusion about how to perform the task. Participants are asked to state first-order beliefs, but are incentivized to match the modal response in their responses. Therefore, they might be confused about which level of beliefs to report. Both of these aspects of the KW method contribute to uncertainty over one's own optimal decision, which has been linked to systematic attenuation of economic decisions and compressing towards intermediate cognitive defaults (Enke 2023).

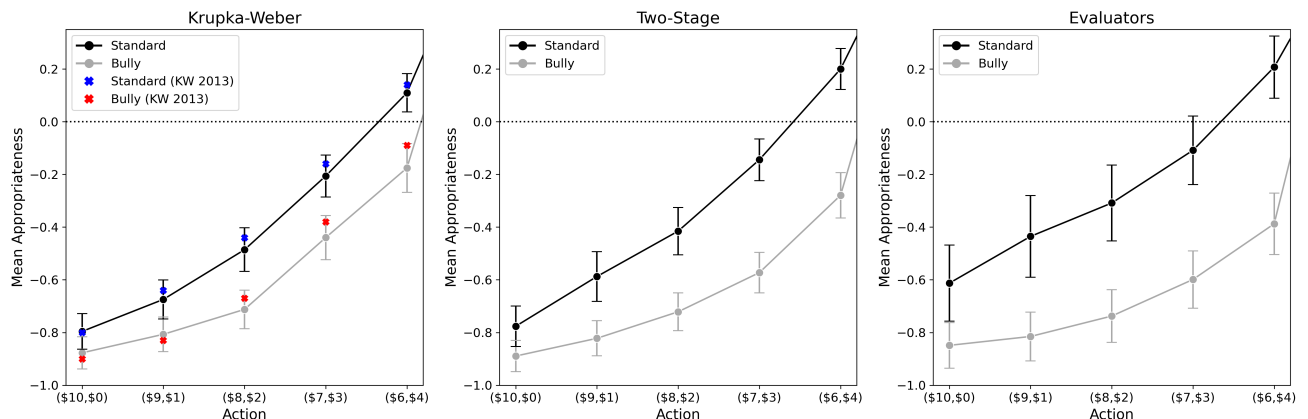


Figure 1: Giving and Taking Norms by Elicitation Method

3.4 Comprehension

To further investigate the role of confusion, we examine the comprehension check, which consists of one multiple choice question asking participants what they should be doing in the subsequent norm elicitation task. Figure 2 reports the estimated probability of passing the comprehension check on the first attempt by treatment, using a probit regression. Almost 50% of participants fail the comprehension check on their initial attempt under the KW methods, as opposed to less than 10% for the evaluators. As anticipated, the

Second-Stage methods are simpler for participants to understand than the KW methods, but more complex than being asked for first-order beliefs about social appropriateness (Evaluators). Poor understanding of the task in the KW methods is consistent with König-Kersting (2021), who finds that including a comprehension check in the pre-task instructions improves post-task recall of the task and incentives, yet has no effect on elicited norms. Our comprehension check requires participants to demonstrate their understanding of the task before starting the task unlike Krupka-Weber (2013). Despite this, Figure 1 shows few differences between the average norm ratings in our KW methods and the original paper.⁶ While there is substantial confusion with the task at hand in KW methods, the severity of the problem is unclear as attempts to reduce dissonance and confusion have not produced meaningful differences in norms.

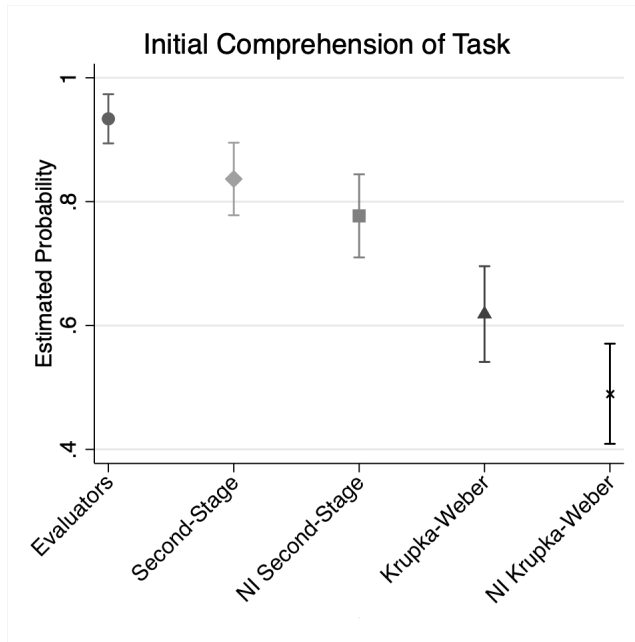


Figure 2: Comprehension Check

4 Conclusion

Our study assesses two dimensions of eliciting social norms: the use of financial incentives and the choice between the coordination game approach of Krupka & Weber (2013), and a two-stage method that directly measures first and second-order beliefs separately. We implement five methods that differ in the type of beliefs they elicit and in the use of monetary incentives. First, we assess these methods on their ability to identify a difference in social norms where we expect it to exist. Using a standard dictator game and a variant with differing initial endowments (“Bully Game”), we replicate KW’s finding of a qualitative difference in norms between these games. All methods show that taking money is less socially appropriate than giving money, holding outcomes fixed, regardless of the presence of monetary incentives. We find that the difference in social appropriateness between the standard and bully dictator games varies across methods, with the Evaluators (first-order beliefs) exhibiting the largest gap in social appropriateness and the KW methods exhibiting the smallest gap. A comprehension check before the norm elicitation reveals that nearly half of KW participants initially misunderstood the task. Our results suggest that complex norm elicitation methods may be less sensitive to the norm-relevant context than measuring first-order or second-order beliefs directly due to attenuation arising largely from strategic uncertainty rather than confusion. Due to this, the KW method provides a conservative test for uncovering differences in social norms between different environments.

Support

The authors are grateful for financial support from the Behavioral Economics Design Initiative at the University of Pittsburgh. Wang completed this work as a Fellow at the Center for Advanced Study in the Behavioral Sciences.

⁶Comparing our *Krupka-Weber* incentivized method to the original data, only 3/22 norms are significantly different at the 0.05 level using rank-sum tests.

References

- [1] Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115-1153.
- [2] Aycinena, D., Rentschler, L., Beranek, B., & Schulz, J.F. (2022). Social norms and dishonesty across societies. *Proceedings of the National Academy of Sciences*, 119(31), e2120138119.
- [3] Barr, A., Lane, T., & Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164, 153-164
- [4] Bicchieri, C., & Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2), 191-208.
- [5] Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132, 59-72.
- [6] Bicchieri, C., Dimant, E., Gelfand, M., & Sonderegger, S. (2023). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior & Organization*, 205, A4-A7.
- [7] Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American economic review*, 110(10), 2997-3029.
- [8] 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.
- [9] Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4), 99-117.
- [10] Enke, B. (2023). Market exposure and human morality. *Nature Human Behaviour*, 7(1), 134-141.
- [11] Fallucchi, F., & Nosenzo, D. (2022). The coordinating power of social norms. *Experimental Economics*, 25(1), 1-25.
- [12] Görge, L., & Nosenzo, D., (2020). Measuring social norms in economics: Why it is important and how it is done. *Analyse & Kritik*, 42(2), 285-312.
- [13] König-Kersting, C. (2024). On the robustness of social norm elicitation. *Journal of the Economic Sciences Association*.
- [14] Krupka, E. L., & 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.

Appendix

Table 4: Instructions for main task in each condition

Treatments	Instructions
<i>Evaluators</i>	<i>For each of the choices, please indicate whether you believe choosing that option is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate.</i>
<i>Second-Stage</i>	<i>For each of the choices, please indicate your guess for the most frequently given by the evaluators. Remember that you will earn money (\$2) if your response to a randomly-selected question is the same as the most common rating provided by the evaluators.</i>
<i>Non-Incentivized Second-Stage</i>	<i>For each of the choices, please indicate your guess for the most frequently given by the evaluators.</i>
<i>Krupka-Weber</i>	<i>For each of the choices, please indicate whether you believe choosing that option is very socially inappropriate, somewhat socially inappropriate, somewhat socially appropriate, or very socially appropriate. To indicate your response, please place a checkmark in the corresponding box. Remember that you will earn money (\$2) if your response to a randomly-selected question is the same as the most common response provided in this study.</i>
<i>Non-Incentivized Krupka-Weber</i>	<i>For each of the choices, please indicate your guess for the most frequently given rating in this study.</i>

Table 5: Effects of Incentivization on Rating Variance

Outcome	Non-Incentivized			Incentivized			F-value
	Obs	Mean	Std Dev	Obs	Mean	Std Dev	
S (10,0)	228	-0.752	0.498	162	-0.759	0.502	0.983
S (9,1)	228	-0.593	0.572	162	-0.604	0.568	1.013
S (8,2)	228	-0.424	0.561	162	-0.429	0.584	0.924
S (7,3)	228	-0.172	0.515	162	-0.153	0.531	0.941
S (6,4)	228	0.162	0.490	162	0.162	0.489	1.005
S (5,5)	228	0.794	0.362	162	0.828	0.376	0.925
S (4,6)	228	0.441	0.558	162	0.513	0.553	1.018
S (3,7)	228	0.309	0.613	162	0.287	0.665	0.847
S (2,8)	228	0.146	0.736	162	0.157	0.762	0.935
S (1,9)	228	0.073	0.791	162	0.099	0.821	0.929
S (0,10)	228	0.051	0.842	162	0.044	0.879	0.918
B (10,0)	218	-0.864	0.381	143	-0.894	0.355	1.155
B (9,1)	218	-0.807	0.413	143	-0.827	0.397	1.080
B (8,2)	218	-0.718	0.439	143	-0.727	0.447	0.964
B (7,3)	218	-0.534	0.494	143	-0.524	0.487	1.028
B (6,4)	218	-0.288	0.541	143	-0.231	0.541	0.997
B (5,5)	218	0.817	0.419	143	0.808	0.412	1.035
B (4,6)	218	0.438	0.588	143	0.497	0.509	1.338*
B (3,7)	218	0.289	0.656	143	0.360	0.608	1.161
B (2,8)	218	0.179	0.715	143	0.281	0.692	1.067
B (1,9)	218	0.099	0.764	143	0.238	0.751	1.021
B (0,10)	218	0.062	0.829	143	0.178	0.817	1.028

This table tests the statistical difference between the variance of ratings of non-incentivized participants and incentivized participants. Outcomes are presented with the moving player first such that the outcome (10,0) indicates the dictator ending up with \$10 and the other player receiving \$0. The "S" and "B" before the outcome represent the Standard and Bully variants of the dictator game, respectively.