ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Fallucchi, Francesco; Nosenzo, Daniele

Working Paper

The coordinating power of social norms

CeDEx Discussion Paper Series, No. 2020-14

Provided in Cooperation with:
The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx)

Suggested Citation: Fallucchi, Francesco; Nosenzo, Daniele (2020) : The coordinating power of social norms, CeDEx Discussion Paper Series, No. 2020-14, The University of Nottingham, Centre for Decision Research and Experimental Economics (CeDEx), Nottingham

This Version is available at:
https://hdl.handle.net/10419/228378

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.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

University of Nottingham
UK | CHINA | MALAYSIA

Discussion Paper No. 2020-14

Francesco Fallucchi and
Daniele Nosenzo

June 2020

The Coordinating Power of Social Norms

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

THE COORDINATING POWER OF SOCIAL NORMS

15 June 2020

Francesco Fallucchi[a] and Daniele Nosenzo[b]

Abstract: A popular empirical technique to measure norms uses coordination games to elicit what subjects in an experiment consider appropriate behavior in a given situation (Krupka and Weber, 2013). The Krupka-Weber method works under the assumption that subjects use their normative expectations to solve the coordination game. However, subjects might use alternative focal points to coordinate, in which case the method may deliver distorted measurements of the social norm. We test the vulnerability of the Krupka-Weber method to the presence of alternative salient focal points. We find that the method is robust as long as there are clear normative expectations about what constitutes appropriate behavior. In settings where there is a less clear consensus about the social norm, the method is more vulnerable.

Keywords: Social Norms; Krupka-Weber method; Coordination; Focal Point; Saliency; Dictator Game.

JEL Classification: C72, C91

Acknowledgements: We received helpful comments from Abigail Barr, Zvonimir Bašić, Georgia Michailidou and Bob Sugden. We acknowledge funding from the Luxembourg Institute of Socio-Economic Research (LISER).

[a] Luxembourg Institute of Socio-Economic Research (LISER), Luxembourg.
[b] Department of Economics and Business Economics, Aarhus University, Denmark & School of Economics, University of Nottingham, UK.

1. INTRODUCTION

The notion of "norms" is increasingly used in economics to explain a variety of behavioral phenomena (e.g., Gächter et al., 2013; Krupka and Weber, 2013; Reuben and Riedl, 2013; Butler et al., 2016; Kimbrough and Vostroknutov, 2016; Bursztyn et al., 2018). A defining characteristic of *social* norms is that they embody *shared expectations* about which actions are appropriate or inappropriate in a particular situation (Ostrom, 2000; Bicchieri, 2006; Schram and Charness, 2015; Fehr and Schurtenberger, 2018; d'Adda et al., 2019).[1] This feature of social norms has been exploited in empirical research to design experiments with incentive-compatible mechanisms to elicit norms. In these experiments subjects receive incentives to guess other people's expectations about the social acceptability of various behaviors in a given situation. These second-order beliefs reveal subjects' perception of the shared normative expectations (and hence the social norm) that prevail in that situation.

One such elicitation mechanism has been proposed by Krupka and Weber (2013). Their method has quickly gained popularity in the discipline: first released as a working paper in 2008, the study has already received nearly 500 citations, along a path of exponential growth (Figure 1). The Krupka-Weber technique has been used to explain a large series of phenomena, such as reciprocity (Gächter et al., 2013; Nikiforakis et al., 2014), fair sharing (Gächter et al., 2017), promise keeping (Krupka et al., 2017), lying (d'Adda et al., 2017), ethical conduct of financial advisers (Burks and Krupka, 2012), corruption (Banerjee, 2016), discrimination (Barr et al., 2018), and gendered occupational choices (Gangadharan et al., 2016).

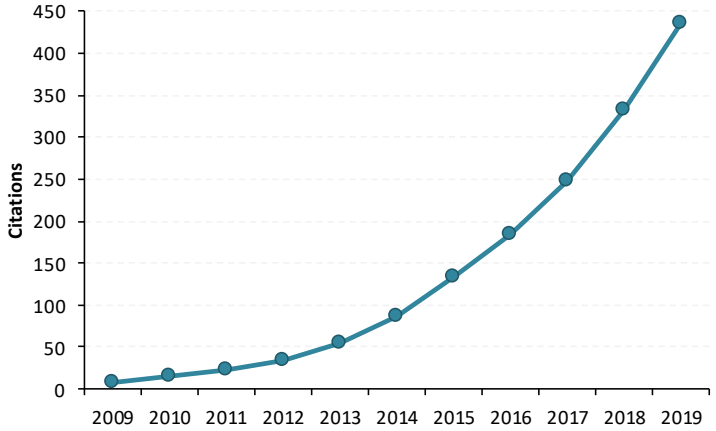
In the Krupka-Weber task, subjects are described a situation where a decision-maker can choose among several actions. Subjects rate each action according to its "social appropriateness" and receive a monetary reward if they rate the actions in the same way as most other participants in the task. The incentive structure transforms the task into a *pure coordination game* where participants have to tacitly coordinate with others in the way they rate actions. If subjects use their shared normative expectations to coordinate in the task, then their ratings will indirectly reveal the social norm that prevails in the situation they are asked to evaluate.

However, the coordination game that subjects play in the Krupka-Weber task has multiple equilibria, all yielding the same payoff to all players. Subjects may in principle coordinate on any

[1] See Bicchieri (2006) for a discussion of the differences between social norms, moral norms and personal norms.

rating for a specific action, and not necessarily on the rating that is consistent with their normative expectations. Thus, the Krupka-Weber mechanism delivers an accurate measurement of norms only under the assumption that the saliency of normative expectations as a coordination device dominates the saliency of any other coordination device that subjects may consider to solve the coordination game. In this paper, we propose an empirical test of this assumption.

Figure 1 – Cumulative citations of Krupka and Weber (2013), 2009-2019



Note: Sourced from Google Scholar (27 May 2020, current citation count = 492). The Figure only displays citations between 2009 and 2019.

The logic of our test is as follows. We construct a version of the Krupka-Weber procedure where we deliberately introduce an extraneous salient feature of the task that subjects can use as a coordination device in alternative to their normative expectations. We compare the ratings that subjects select in this modified task to those elicited in the standard Krupka-Weber task in order to assess the distortions caused by the presence of alternative focal points that are unrelated to normative expectations.

We introduce the alternative focal point by attaching *visual labels* (smileys or frowny faces) to the social appropriateness ratings that subjects use in the Krupka-Weber task. We vary the size of one of the smileys to make it stand out compared to the others (see Table 1). In two pilot experiments, we test the saliency of the visual labels by checking whether: (i) the oversized smiley is recognized as an "oddity", i.e. an element that stands out relative to others, and (ii) the oddity creates a focal point in a pure coordination game where subjects have to choose the same visual label as most other participants. We find that a vast majority of subjects recognize the

oversized smiley as an oddity and that around 80% choose it to coordinate with others in the coordination game. The pilot results confirm that our visual labels constitute a salient focal point.

We then conduct a norm-elicitation experiment using the Krupka-Weber task. In a baseline treatment, we use the standard version of the task, without visual labels. In two treatments we introduce the visual labels (either with or without an oversized smiley). In all treatments we elicit the norm of giving in a standard dictator game. We collect data from a large sample (N = 773) in order to have sufficient power to detect small effects.

We chose the dictator game for two reasons. First, several previous experiments have elicited norms in this game (Krupka and Weber, 2013; Erkut et al., 2015; Kimbrough and Vostroknutov, 2016; 2018; Chang et al., 2019), providing a clear benchmark for the normative expectations that prevail in this setting. Second, we know from these experiments that an interesting feature of the norm of giving in dictator games is that, while subjects seem to have clear normative expectations about the social appropriateness of giving little to the recipient (a large majority of people find it unacceptable), there is much less consensus on the appropriateness of overgenerous actions that leave more than the equal split to the recipient. This is useful because it allows us to test, in the context of the same game, the robustness of the Krupka-Weber method both in a situation where there are clear normative expectations about the behavior being evaluated and in one where normative expectations are more heterogeneous. The saliency of normative expectations in the Krupka-Weber task may be stronger in the former case than in the latter, which may have implications for how vulnerable the method is to the presence of alternative focal points.

We find that the Krupka-Weber method is broadly robust to the presence of alternative focal points. Dictator game actions are rated similarly across our three treatments. There is, however, a key distinction between actions that are regulated by a clear norm (giving less or equal to the equal split) and actions for which there is a less clear consensus about whether they are socially appropriate or inappropriate (giving more than the equal split). In the former case, despite the high power of our tests to detect small treatment effects, we see no distortion in the measurement of normative expectations across our treatments. In the latter case, we find instead a small distortion: a significantly larger proportion of subjects seem to use the alternative focal point to coordinate in the tasks with visual labels compared to the standard Krupka-Weber task.

Our paper contributes to the literature on the empirical measurement of social norms. We show that, if a clear social norm exists in a particular situation, then the Krupka-Weber method can reliably elicit it. However, if there is a lack of consensus about what people view as socially appropriate, the method is more vulnerable to measurement distortions that are caused by its incentive structure: instead of their normative expectations, subjects may use alternative focal points that are available in the task to coordinate. Our recommendation to researchers using the Krupka-Weber method is to be cautious in drawing inferences from situations where normative expectations are divergent. Additional robustness checks may be needed in these cases in order to exclude that alternative focal points have confounded the elicitation of normative expectations.

More broadly, our paper also contributes to the literature on focal points in coordination games. Label-salient focal points have been shown to be very effective solutions to coordination problems (e.g., Mehta et al., 1994a; Mehta et al., 1994b; Bacharach and Bernasconi, 1997), especially when coordination yields symmetric payoffs to all players (Crawford et al., 2008). We compare the coordination power of a label-salient focal point with that of normative expectations in a pure coordination game that is framed as a norm-elicitation task (the Krupka-Weber task). We show that normative expectations dominate the alternative label-salient focal point. Our results illustrate the power of social norms to coordinate beliefs and actions in social situations.

2. EXPERIMENTS

We ran three different types of experiments. In our main experiment, we used the Krupka-Weber technique (Krupka and Weber, 2013) to elicit social norms in a dictator game. In the standard Krupka-Weber elicitation task, subjects are asked to evaluate the social appropriateness of the actions available to the dictator. Subjects use a 4-point scale to rate actions: very inappropriate; somewhat inappropriate; somewhat appropriate; or very appropriate. In the main experiment, we will compare the standard version of the Krupka-Weber task with two modified versions of the task where we associate visual labels to the 4 appropriateness ratings available to the subjects. Before the main experiment, we conducted two pilot experiments to choose to optimal way to display these visual labels and calibrate the design of the main experiment. In this section, we first describe the design and results of the pilot experiments. We next describe the design of the main experiment. We report the results of the main experiment in the next section.

2.1 Pilot experiments: Design

The visual labels we used in our study are the 4 smiley faces that are often used in surveys and public spaces (e.g. airports) to collect feedback on users' satisfaction levels (Table 1). In the main experiment, we associate each smiley to an item of the 4-point scale in the Krupka-Weber task. Our aim is to use the visual labels to introduce an *additional focal point* in the Krupka-Weber task that subjects may use as an alternative to normative expectations to coordinate in the norm-elicitation experiment. This will allow us to test the robustness of the Krupka-Weber method when alternative focal points are available to subjects.

For our test to work, however, it is crucial that the visual labels introduce a *salient* focal point in the coordination game, so that subjects may be tempted to use the labels, instead of normative expectations, to coordinate in the Krupka-Weber task. We conjectured that the visual labels may introduce a salient focal point if one of the 4 smiley faces would somehow "stand out" relative to the others. We therefore designed various versions of the labels, where we gradually increased the size of one of the 4 smileys to make it stand out (see Table 1). We designed two pilot experiments to test the extent to which the different versions of the visual labels succeeded in creating a salient focal point. Inspired by Bacharach and Bernasconi (1997), we measure the saliency of a focal point by eliciting (i) subjects' propensity to notice an object as an "oddity" that stands out relative to the other objects available, and (ii) subjects' propensity to use the oddity to solve an actual pure matching coordination game (see also Mehta et al., 1994a; Mehta et al., 1994b; Bardsley et al., 2010; Hargreaves Heap et al., 2014; Charness and Sontuoso, 2019).

In our first pilot experiment, we recruited 251 participants on Amazon Mechanical Turk (AMT) and randomly showed them 1 out of the 5 possible versions of the 4 smiley faces displayed in Table 1.[2] In a first treatment (T1), all 4 objects had the same size (80 pixels). In the other four treatments the size of the right-most object (the dark green smiley) was increased to 85 pixels (T2), 90 pixels (T3), 95 pixels (T4), and 100 pixels (T5). Each subject was assigned to one treatment only. In all treatments, we asked subjects to indicate which of the four objects, in their opinion, "stood out the most". The question was unincentivized (subjects received a $0.50 fixed fee as reward for participating). After subjects had expressed their opinion, the experiment ended

[2] All experiments were programmed using the software LIONESS (Giamattei et al., 2019).

with a short questionnaire collecting basic demographic information and free-form comments about the task.

Our second pilot experiment uses the same five treatments displayed in Table 1, but in the context of a pure coordination game. We recruited from AMT another 252 subjects (who had not participated in the first pilot) and randomly assigned them to 1 of the 5 treatments. Subjects were asked to choose one of the 4 smiley faces, knowing that if they chose the same smiley face as most other participants, they would be paid $2 (in addition to a $0.50 fixed participation fee); if they did not choose the most frequently-chosen smiley face, they were only paid the fixed participation fee.[3] This incentive system mimics the incentivization that is commonly used in the Krupka-Weber method and that we will also use in the main experiment (see section 2.3). Subjects made their choice only once and then the experiment ended with the same brief questionnaire as in the first pilot.

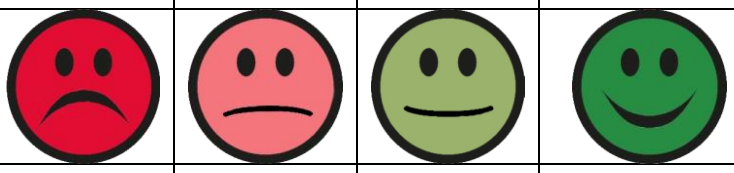
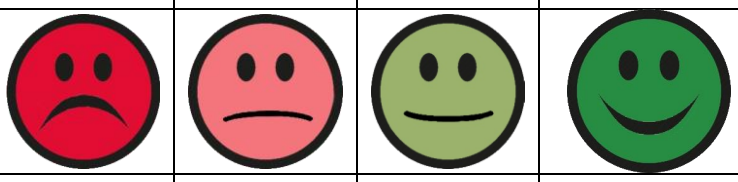
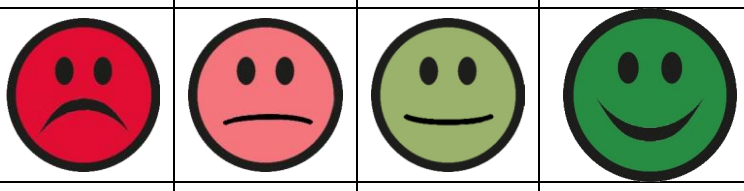
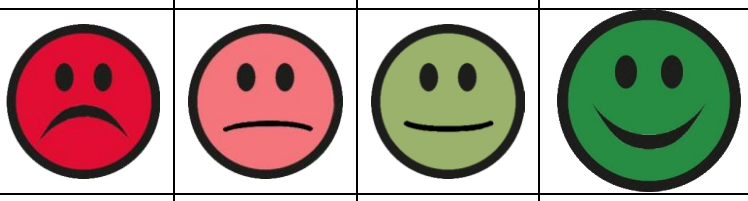
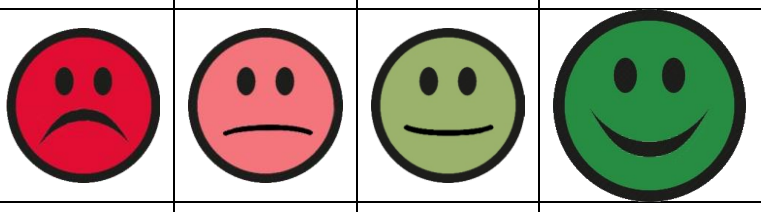
2.2 Pilot experiments: Results

Table 1 shows the results of the two pilot experiments. In the first row below each treatment, we report the percentage of subjects from the first pilot who indicated that a particular object "stands out the most". When the four objects have all the same size (T1), the two end-point objects are those that stand out the most, particularly the left-most, dark red smiley (42% indicated it stands out). In the other four treatments, as the size of the right-most smiley increases, an increasingly larger share of subjects recognizes the oversized object as an oddity. The percentage of subjects selecting the dark green smiley exceeds 80% in T4 and T5. A series of Fisher's exact tests indicate that the distribution of responses in T4 and T5 is significantly different than in T1, T2 and T3 (all $p \leq 0.030$).[4] These results indicate that a difference of 15 pixels is sufficient to lead a large majority of subjects to recognize the oversized object as an oddity.

[3] In the instructions we explicitly told subjects that the other participants they were trying to coordinate with were other AMT workers from the US who were completing the same task, using the same decision screen as the subject. See Appendix A in the Online Supplemental Materials (OSM) for the instructions used in the study.

[4] We find no difference between T4 and T5 ($p = 0.489$), or between T1 and T2 ($p = 0.253$). We find significant differences between T3 and both T1 and T2 ($p = 0.012$ in either case). Throughout the paper we report p-values corrected for multiple comparisons using the Benjamini and Hochberg (1995) false discovery rate procedure. In the case of the two pilot experiments, we correct for the fact that there are 10 possible bilateral comparisons between the 5 treatments.

Table 1 - Pilot experiments: Results

Treatments				
T1 (80px)				
% saying the object "stands out" (N=50)	42%	14%	12%	32%
% selecting the object in coordination game (N=50)	2%	2%	18%	78%
T2 (85px)				
% saying the object "stands out" (N=50)	42%	4%	8%	46%
% selecting the object in coordination game (N=52)	2%	2%	27%	69%
T3 (90px)				
% saying the object "stands out" (N=50)	14%	16%	8%	62%
% selecting the object in coordination game (N=51)	0%	6%	16%	78%
T4 (95px)				
% saying the object "stands out" (N=50)	12%	2%	2%	84%
% selecting the object in coordination game (N=50)	0%	2%	10%	88%
T5 (100px)				
% thinking the object "stands out" (N=51)	10%	0%	8%	82%
% selecting the object in coordination game (N=49)	0%	0%	20%	80%

The second row below each treatment shows the percentage of subjects who chose a specific object in the coordination game. The right-most smiley is chosen very frequently already in T1, where all 4 objects have the same size (78% of subjects chose it). Free-form comments collected in the post-experimental questionnaire suggest that this is partly because subjects associate the happy face of the object with the monetary reward at stake in the game and they believe that this makes it a natural choice to coordinate with the co-players. Increasing the size of the dark green smiley by 15 pixels or more (T4 and T5) increases the rate of subjects who choose it in the coordination game to above 80%, although the differences relative to the other treatments are insignificant (Fisher's exact tests, $p \geq 0.430$).

Overall, the two pilots indicate that the smiley labels can successfully create a focal point of coordination, especially when one of the smileys is oversized relative to the others. Among the versions we tested, those in T4 or T5 with an oversized object of 95 or 100 pixels, introduce particularly salient focal points: more than 80% of subjects recognize the oversized object as an oddity that is clearly different from the others, and more than 80% of subjects choose the oddity in an actual coordination game.

In our main experiment, described in the next sub-section, we will use the visual labels of T1 and T4 to introduce an alternative, salient focal point in the Krupka-Weber norm-elicitation task. The two treatments differ in the "strength" of the alternative focal point: the results of the pilots indicate that the saliency of the dark green smiley is higher in T4 since in this treatment the oversized object was selected by the highest share of subjects in both pilots. In contrast, in T1 subjects did not recognize the object as an oddity, although they used it to solve the coordination game of the second pilot.

2.3 Main Experiment

In the main experiment, we used the Krupka-Weber method to measure social norms in a standard dictator game where the dictator allocates $10 between themselves and a passive recipient. We chose to focus on the dictator game because the norms that pertain this game have already been measured in several experiments across multiple subject pools, yielding a broadly consistent picture of what people believe to be appropriate behavior in this setting (Krupka and Weber, 2013; Erkut et al., 2015; Kimbrough and Vostroknutov, 2016; 2018; Chang et al., 2019). In previous experiments, the most appropriate action is for the dictator to split equally the

endowment. Actions that leave the recipient with less than half of the endowment are generally viewed as less appropriate, and giving nothing to the recipient is most inappropriate. Actions that leave the recipient with more than half of the endowment are generally viewed as appropriate.

However, there is an important distinction between actions that lie on either side of the equal split. While there is typically a strong consensus that giving less than the equal split is inappropriate (a clear majority rate these actions this way), there is considerably more heterogeneity in regards to actions that leave recipients with more than the equal split. That is, while people seem to recognize the existence of a clear norm that stigmatizes dictators for keeping more than half of the endowment, it is less clear that there is a shared understanding of how to evaluate overgenerous actions. This is useful for our purposes because in the context of the same game we can assess the robustness of the Krupka-Weber technique both when there are clear shared expectations about how to evaluate actions (in which case we would expect normative expectations to constitute a salient focal point), and when normative expectations are more heterogeneous (in which case they may not constitute a salient focal point). Thus, our results will be informative about the potential distortions that may occur when a researcher is trying to measure normative judgments in settings where a clear norm exists and in settings where it does not.

Our experiment consists of three treatments. In the baseline condition (**KW**), we implement the standard Krupka-Weber method to elicit social norms. We recruited 261 AMT workers form the US and described them a dictator game where the dictator can divide $10 between themselves and an anonymous recipient in increments of $1. We asked them to imagine that dictator and recipient were both US subjects recruited on AMT. For each of the 11 possible divisions of the $10, subjects indicated whether they thought the allocation was "very socially inappropriate", "somewhat socially inappropriate", "somewhat socially appropriate", or "very socially appropriate". Subjects were told that "socially appropriate" means "*behavior that you think most people would agree it is the correct thing to*". Subjects were paid based on how they rated the 11 actions they had to evaluate. At the end of the experiment, we randomly selected one of the 11 actions and compared a subject's evaluation of that action with the way that most other participants had evaluated the same action (we told subjects that these other participants were also AMT workers recruited in the US). If their evaluation matched that of most other

participants, subjects were paid an additional $2 on top of their $0.50 fixed participation fee; otherwise they were only paid the participation fee. After the evaluation of the 11 actions, subjects completed a short questionnaire collecting basic demographic information and free-form comments on the task, and this concluded the experiment.

In the other two treatments, we asked subjects to evaluate the same 11-action dictator game, keeping instructions and procedures identical to the baseline condition. The only difference was that we associated each item of the appropriateness scale with a smiley face: "frowny" faces for inappropriate ratings and "smiley" faces for appropriate ratings (Figure 2). In a first treatment (**SMILE-80px**), we used 4 smiley faces all of the same size (80 pixels, this corresponds to treatment T1 in the pilot experiments). In the second treatment (**SMILE-95px**), we used the smiley faces of T4 in the pilot experiments, where the dark green smiley was 95 pixels in size, while the other three smileys were 80 pixels.

Figure 2 - Appropriateness scale used in SMILE-80px

Very socially inappropriate | Somewhat socially Inappropriate | Somewhat socially appropriate | Very socially appropriate

As discussed in section 2.2, in SIMILE-95px both measures of saliency (propensity to recognize the oversized object as an oddity and propensity to choose it in a coordination game) are aligned, while in SIMILE-80px the two measures are not aligned: while most subjects chose the dark green smiley in the coordination game, only 32% thought that it "stands out" relative to the other objects. We can therefore use the two treatments to test the robustness of the Krupka-Weber method to the inclusion of alternative focal point that are more (SIMILE-95px) or less (SIMILE-80px) salient.

We recruited 258 AMT workers for the SIMILE-80px treatment and another 254 workers for the SIMILE-95px treatment. Thus, the total sample size of the main experiment is of 773 AMT workers. Each subject participated in one treatment only. The sample size was determined before running the experiment using power analysis in order to be able to detect effects of size

0.25 or larger (Cohen's d) with 80% power and $\alpha = 0.05$, based on a two-tailed Mann-Whitney rank-sum test. This is an important feature of our experiment since it allows us to perform tests that have sufficient power to detect even small distortions in the Krupka-Weber ratings.

3. Results of the Main Experiment

Following the literature, we assign evenly-spaced numeric values to the four appropriateness ratings: -1 to "very socially inappropriate", -0.33 to "somewhat socially inappropriate", +0.33 to "somewhat socially appropriate" and +1 to "very socially appropriate". Figure 3 shows, for each treatment, the average rating of the 11 actions in the dictator game. Table 2 shows the distribution of ratings across the three treatments.

In all treatments, the most appropriate action is to share the $10 equally with the recipient; the least appropriate action is to give $0 to the recipient. Leaving less than $4 to the recipient is considered inappropriate, while giving more than the equal split is generally viewed as appropriate. However, there is weaker consensus about how to rate these overgenerous actions, as we can see from our standard replication of the Krupka-Weber method (KW treatment). These results are fully in line with the findings of previous experiments that elicited norms in the dictator game.

Generally, actions are rated similarly across the three treatments. Figure 3 shows that the differences between treatments emerge more markedly for actions that leave the recipient with more than $5. These actions are evaluated as relatively more appropriate in the two treatments with visual labels (SMILE-80px and SMILE-95px) than in the KW baseline. Actions that leave the recipient with $5 or less are rated instead very similarly across all treatments.

We test for treatment differences in two ways. First, we use Mann-Whitney rank-sum tests to compare, for each of the 11 actions, the distributions of ratings between the KW baseline condition and each of the two treatments with visual labels. Second, we test the more specific hypothesis that the treatments with visual labels generated a larger share of "very socially appropriate" ratings compared to the KW baseline - since this was the rating associated with the salient smiley face. For this second test we use Fisher's exact tests. Because we perform a total of 44 tests (11 actions compared between KW and each of the two treatments using two types of tests), the p-values reported below are corrected for multiple comparisons using the false

discovery rate procedure. We use a type I error rate of 5% as the threshold for statistical significance.

Figure 3 - Mean appropriateness ratings

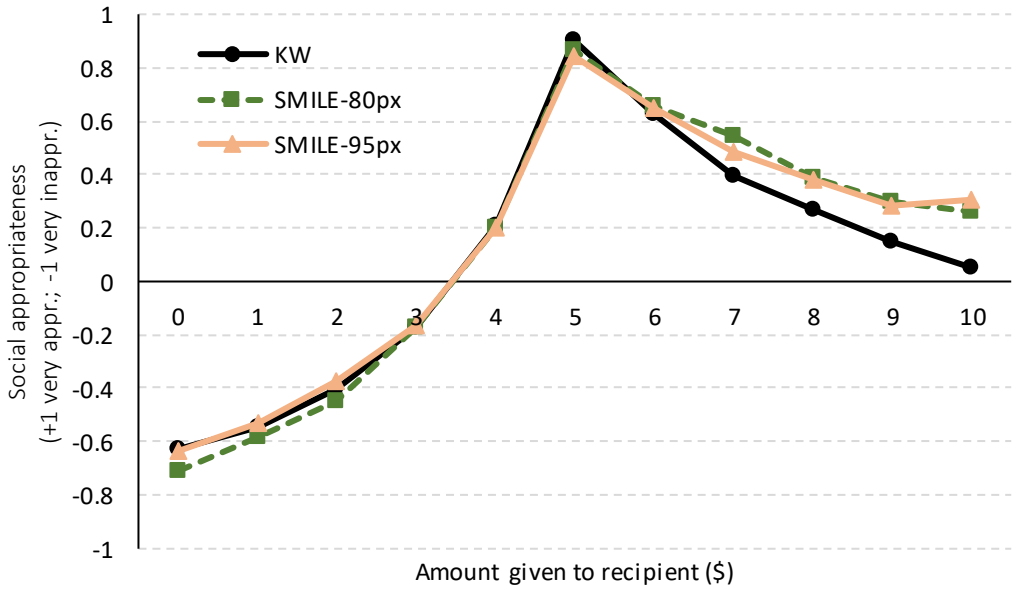


Table 2 - Distribution of appropriateness ratings

Amount given	KW (N=261)				SMILE-80px (N=258)				SMILE-95px (N=254)			
	-1	**-0.33**	**+0.33**	**+1**	**-1**	**-0.33**	**+0.33**	**+1**	**-1**	**-0.33**	**+0.33**	**+1**
0	62	23	10	4	73	14	10	3	65	21	9	5
1	54	31	9	6	59	26	10	5	50	32	14	4
2	38	42	13	7	40	42	15	4	34	45	15	6
3	14	56	23	7	12	59	22	7	12	58	23	7
4	2	29	56	14	0	31	57	12	2	28	59	12
5	0	0	13	86	0	1	18	81	0	2	17	80
6	1	9	34	56	1	7	34	58	2	6	36	57
7	5	20	35	40	3	17	26	54	4	19	28	49
8	13	22	27	38	13	20	13	54	13	20	14	53
9	23	18	22	37	23	10	14	52	20	15	15	49
10	34	11	18	37	28	9	10	53	23	9	15	52

Note: Each cell shows the percentage of subject selecting a specific appropriateness rating for a given action (amount given in the dictator game). The grey shaded cells indicate the modal response for each given action -1 = very socially inappropriate; -0.33 = somewhat socially inappropriate; +0.33 = somewhat socially appropriate; +1 = very socially appropriate.

Starting with Mann-Whitney rank-sum tests, we only find significant differences between the KW baseline and the treatments in four cases. When the dictator gives \$7 or \$9 to the recipient we find significant differences between KW and SMILE-80px ($p = 0.012$ for \$7; $p = 0.046$ for \$9). Moreover, when the dictator gives \$10 to the recipient we find significant differences between KW and both SMILE-80px and SMILE-95px ($p = 0.017$ and $p = 0.006$, respectively). In all other cases, the differences are not significant at the 5% level.

The results of the Fisher's exact tests are similar. We find significant differences between KW and either of the two treatments whenever the dictator gives \$8 or more to the recipient (all $p \leq 0.006$). We also find a significant difference between KW and SMILE-80px when the dictator gives \$7 ($p = 0.006$). In all these cases, Table 2 shows that the share of subjects who select the rating "very socially appropriate" for the action increases by at least 12 percentage points in the treatments compared to the KW baseline. In all other cases, the differences in the share of the "very socially appropriate" rating are much smaller and statistically insignificant.

4. Discussion & Conclusions

Our paper contributes to the empirical literature on the measurement of social norms. The popular method to elicit norms introduced by Krupka and Weber (2013) suffers from a theoretical weakness. The method elicits social norms using a coordination game where subjects are incentivized to evaluate behavior in the same way as most other subjects. Because the instructions for the task are heavily framed in the language of social norms, it is plausible that subjects may use normative expectations (i.e. their beliefs about what others think is socially appropriate) as a coordination device to match other participants' evaluations. However, in principle, subjects may use *any* other salient feature of the task to coordinate, in which case the elicitation technique will return a distorted measurement of the social norm. Our contribution is to propose an empirical test of the vulnerability of the Krupka-Weber technique to the presence of focal points that are not related to the norm.

Our results are clear. In settings where there are convergent normative expectations about what is appropriate or inappropriate behavior, the Krupka-Weber method is robust to the presence of alternative salient focal points. In the dictator game we studied in our experiment, there is a clear consensus about how to evaluate actions that leave the recipient with an equal

split of money or less. A large majority of participants find the equal split very appropriate, while most participants think that giving the recipient less than the equal split is inappropriate. This is true in our experiment as well as in previous studies eliciting norms in the dictator game (Krupka and Weber, 2013; Erkut et al., 2015; Kimbrough and Vostroknutov, 2016; 2018; Chang et al., 2019). For these actions, we find that the presence of an alternative focal point that is salient and is disconnected from the norm has no discernible effect on the measurement of norms. In interpreting this finding, we emphasize that our study is highly powered (we can detect effects of size d = 0.25 or larger with 80% power), and so this null result is not due to lack of power to detect small effects. Rather, we interpret the null result as an indication that, when there are clear normative prescriptions about the behavior under evaluation, subjects use predominantly these prescriptions to solve the coordination game that is inherent to the Krupka-Weber task. This is good news for researchers that intend to use the Krupka-Weber method in their empirical research: if a clear norm exists for the behavior that a researcher aims to study, the Krupka-Weber elicitation task is likely to measure it accurately.

However, our experiment also shows that the Krupka-Weber task is more vulnerable to the presence of alternative focal points in settings where the social norm is less clear. In the dictator game, actions that leave the recipient with more than half of the amount to be divided are on average evaluated as appropriate, but there is considerable dispersion in subjects' evaluations, with many subjects rating those actions as socially inappropriate. This is a common finding in the literature on social norms in dictator games. In our experiment, we find that the norm measurements for these actions are distorted when we introduce an alternative focal point that is disconnected from the norm. This is less good news for researchers that use the Krupka-Weber method to elicit norms: if there is no clear norm in the situation being evaluated, the Krupka-Weber measurements may be influenced by extraneous but salient focal points that the researcher may inadvertently introduce in the coordination game (e.g., the way the rating scale is displayed on a computer screen; the labels used for the ratings; etc.).

This result warrants caution in interpreting normative evaluations that are dispersed and lack consensus.[5] In these situations, researchers may want to probe the robustness of the

[5] In theory, the presence of alternative focal points may confound the elicitation of norms *without* resulting in dispersed ratings, for instance if the alternative focal point is sufficiently strong to sway a large majority of subjects

elicitation method. One possibility would be to use alternative norm elicitation methods that do not rely on coordination games to elicit social norms, such as for instance the sequential elicitation method used by Bicchieri et al. (2019). Alternatively, researchers could use the design we proposed in our study as a *diagnostic tool* to gauge the reliability of the appropriateness judgments elicited with the Krupka-Weber method. The measurements could be repeated using the smiley labels as a robustness check. If the elicited judgments genuinely reflect underlying normative expectations, the inclusion of the smiley labels in the task should only affect responses minimally. Substantial changes in responses should be taken as an indication that clear normative expectations may not exist in the situation under consideration.

The overall conclusion from our study is that the Krupka-Weber elicitation procedure appears to be a reliable method to elicit social norms in situations where there is a clear consensus about what constitutes appropriate or inappropriate behavior. The method may however be less reliable for situations where there is ambiguity about how to evaluate behavior. Our recommendation is that researchers are cautious in drawing inferences from the Krupka-Weber measurements when there is evidence that normative judgments are heterogeneous and no clear consensus emerges about what constitutes appropriate behavior.

At a more general level, our paper also contributes to the literature exploring the determinants of behavior in coordination problems. Several papers have shown that players often use payoff-irrelevant features of the decision problem to select a "focal" equilibrium, such as the labels attached to individual strategies, the spatial positioning of the objects of choice, or information about players' identity (e.g., Mehta et al., 1994a; Mehta et al., 1994b; Bacharach and Bernasconi, 1997; Holm, 2000; Crawford et al., 2008; Parravano and Poulsen, 2015; Isoni et al., 2013; 2019; Sitzia and Zheng, 2019; He and Wu, 2020). Moreover, scholars have argued that expectations about conventional ways to solve a coordination problem can also constitute salient focal points, and so can *normative expectations* about how players ought to behave (Schelling, 1960; Lewis, 1969; Sugden, 1986; Guala and Mittone, 2010). In our experiment, we expose players to two different types of focal points - one stemming from the way strategies are labeled and the other from normative expectations. By pitting the two focal points against one another,

to use it as a coordination device in the Krupka-Weber task. Empirically, we do not think that this is very plausible given that in our experiment we still observe dispersion in the Krupka-Weber ratings for the overgenerous actions even when we introduce very strong alternative focal points in the task.

we can examine the relative strength of these alternative cues for coordination. We find that normative expectations dominate labeling cues in our decision task.

This result may reflect the strong framing on instructions in the Krupka-Weber task, where the notions of norms and social appropriateness are pervasive. Yet, studies have shown that merely alerting individuals to the existence of a rule, or subtly reminding them about a norm, can trigger substantial behavioral compliance (Krupka and Weber, 2009; Kimbrough and Vostroknutov, 2016; Gächter et al., 2020). Taken together, these results illustrate the power of norms to coordinate human behavior.

REFERENCES

d'Adda, G., D. Darai, N. Pavanini, and R.A. Weber. 2017. "Do Leaders Affect Ethical Conduct?" *Journal of the European Economic Association* 15(6): 1177–1213.

d'Adda, G., M. Dufwenberg, F. Passarelli, and G. Tabellini. 2019. "Partial Norms." . CESifo Working Paper 7568.

Bacharach, M., and M. Bernasconi. 1997. "The Variable Frame Theory of Focal Points: An Experimental Study." *Games and Economic Behavior* 19(1): 1–45.

Banerjee, R. 2016. "On the Interpretation of Bribery in a Laboratory Corruption Game: Moral Frames and Social Norms." *Experimental Economics* 19(1): 240–67.

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden. 2010. "Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning*." *The Economic Journal* 120(543): 40–79.

Barr, A., T. Lane, and D. Nosenzo. 2018. "On the Social Inappropriateness of Discrimination." *Journal of Public Economics* 164: 153–64.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300.

Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bicchieri, C., E. Dimant, S. Gaechter, and D. Nosenzo. 2019. "Observability, Social Proximity, and the Erosion of Norm Compliance." Rochester, NY. SSRN Discussion Paper 3355028.

Burks, S.V., and E. Krupka. 2012. "A Multimethod Approach to Identifying Norms and Normative Expectations Within a Corporate Hierarchy: Evidence from the Financial Services Industry." *Management Science* 58(1): 203–17.

Bursztyn, L., A. González, and D. Yanagizawa-Drott. 2018. "Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia." . NBER Working Paper No. 24736.

Butler, J., P. Giuliano, and L. Guiso. 2016. "Trust and Cheating." *The Economic Journal* 126(595): 1703–38.

Chang, D., R. Chen, and E. Krupka. 2019. "Rhetoric Matters: A Social Norms Explanation for the Anomaly of Framing." *Games and Economic Behavior* 116: 158–78.

Charness, G., and A. Sontuoso. 2019. "The Doors of Perception." . Working paper available at: https://digitalcommons.chapman.edu/esi_working_papers/292/.

Crawford, V.P., U. Gneezy, and Y. Rottenstreich. 2008. "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures." *American Economic Review* 98(4): 1443–58.

Erkut, H., D. Nosenzo, and M. Sefton. 2015. "Identifying Social Norms Using Coordination Games: Spectators vs. Stakeholders." *Economics Letters* 130: 28–31.

Fehr, E., and I. Schurtenberger. 2018. "Normative Foundations of Human Cooperation." *Nature Human Behaviour* 2(7): 458–68.

Gächter, S., L. Gerhards, and D. Nosenzo. 2017. "The Importance of Peers for Compliance with Norms of Fair Sharing." *European Economic Review* 97(C): 72–86.

Gächter, S., L. Molleman, and D. Nosenzo. 2020. "The Behavioral Logic of Rule-Following and Social Norm Compliance." . Manuscript in preparation.

Gächter, S., D. Nosenzo, and M. Sefton. 2013. "Peer Effects in Pro-Social Behavior: Social Norms or Social Preferences?" *Journal of the European Economic Association* 11(3): 548–573.

Gangadharan, L., T. Jain, P. Maitra, and J. Vecci. 2016. "Social Identity and Governance: The Behavioral Response to Female Leaders." *European Economic Review* 90: 302–25.

Giamattei, M., L. Molleman, K. Seyed Yahosseini, and S. Gächter. 2019. "LIONESS Lab–a Free Web-Based Platform for Conducting Interactive Experiments Online." . SSRN working paper N. 3329384.

Guala, F., and L. Mittone. 2010. "How History and Convention Create Norms: An Experimental Study." *Journal of Economic Psychology* 31(4): 749–56.

Hargreaves Heap, S., D.R. Arjona, and R. Sugden. 2014. "How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games With Non-Neutral Frames." *Econometrica* 82(3): 1133–51.

He, S., and J. Wu. 2020. "Compromise and Coordination: An Experimental Study." *Games and Economic Behavior* 119: 216–33.

Holm, H.J. 2000. "Gender-Based Focal Points." *Games and Economic Behavior* 32(2): 292–314.

Isoni, A., A. Poulsen, R. Sugden, and K. Tsutsui. 2013. "Focal Points in Tacit Bargaining Problems: Experimental Evidence." *European Economic Review* 59: 167–88.

Isoni, A., A. Poulsen, R. Sugden, and K. Tsutsui. 2019. "Focal Points and Payoff Information in Tacit Bargaining." *Games and Economic Behavior* 114: 193–214.

Kimbrough, E.O., and A. Vostroknutov. 2016. "Norms Make Preferences Social." *Journal of the European Economic Association* 14: 608–38.

Kimbrough, E.O., and A. Vostroknutov. 2018. "A Portable Method of Eliciting Respect for Social Norms." *Economics Letters* 168: 147–50.

Krupka, E., S. Leider, and M. Jiang. 2017. "A Meeting of the Minds: Informal Agreements and Social Norms." *Management Science* 63(6): 1708–1729.

Krupka, E., and R.A. Weber. 2009. "The Focusing and Informational Effects of Norms on Pro-Social Behavior." *Journal of Economic Psychology* 30(3): 307–20.

Krupka, E., and R.A. Weber. 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11(3): 495–524.

Lewis, D. 1969. *Convention: A Philosophical Study*. Oxford: Blackwell.

Mehta, J., C. Starmer, and R. Sugden. 1994a. "Focal Points in Pure Coordination Games - an Experimental Investigation." *Theory and Decision* 36(2): 163–85.

Mehta, J., C. Starmer, and R. Sugden. 1994b. "The Nature of Salience - an Experimental Investigation of Pure Coordination Games." *American Economic Review* 84(3): 658–73.

Nikiforakis, N., J. Oechssler, and A. Shah. 2014. "Hierarchy, Coercion, and Exploitation: An Experimental Analysis." *Journal of Economic Behavior & Organization* 97: 155–68.

Ostrom, E. 2000. "Collective Action and the Evolution of Social Norms." *The Journal of Economic Perspectives* 14(3): 137–58.

Parravano, M., and O. Poulsen. 2015. "Stake Size and the Power of Focal Points in Coordination Games: Experimental Evidence." *Games and Economic Behavior* 94: 191–99.

Reuben, E., and A. Riedl. 2013. "Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations." *Games and Economic Behavior* 77(1): 122–37.

Schelling, T. 1960. *The Strategy of Conflict*. Cambridge: Harvard University Press.

Schram, A., and G. Charness. 2015. "Inducing Social Norms in Laboratory Allocation Choices." *Management Science* 61(7): 1531–46.

Sitzia, S., and J. Zheng. 2019. "Group Behaviour in Tacit Coordination Games with Focal Points – an Experimental Investigation." *Games and Economic Behavior* 117: 461–78.
Sugden, R. 1986. *The Economics of Rights, Co-Operation and Welfare*. 2005 edition. New York: Palgrave Macmillan.

ONLINE SUPPLEMENTAL MATERIAL FOR THE PAPER

"THE COORDINATING POWER OF SOCIAL NORMS"

BY

FRANCESCO FALLUCCHI AND DANIELE NOSENZO

APPENDIX A – Experimental Instructions

This appendix contains the instructions displayed to subjects on their computer screens in the three experiments. For the two pilot experiments, we report the instructions used in T1. The other treatments only differed in the size of the right-most visual label, as shown in Table 1 in the main text. For the main experiment we indicate which parts of the instructions differed across treatments.

Pilot experiment one – T1

#1 Welcome screen

Welcome

Thank you for participating in our HIT.
Including the time for reading these instructions, the HIT will take about 2 minutes to complete.

During the HIT, please do not close this window or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter and we will not be able to pay you!

For participating you will earn a guaranteed fee of $0.50.
You will receive a code to collect your payment via MTurk upon completion.

#2 Instructions

Instructions

In the next screen we will show you four objects. We ask you to choose the one that in your opinion **stands out the most**.

You will receive **$0.50** for answering this question, irrespective of your answer.

#3 Decision screen

Please select the object that in your opinion stands out the most

Pilot experiment two –T1

#1 Welcome screen

Welcome

Thank you for participating in our HIT.
Including the time for reading these instructions, the HIT will take about 2 minutes to complete.

During the HIT, please do not close this window or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter and we will not be able to pay you!

On top of your guaranteed participation fee of $0.50, you can earn a bonus of $2.00 dependent on your responses and the responses of other participants in this HIT.

You will receive a code to collect your payment via MTurk upon completion.

#2 Instructions

Instructions

In this HIT you will have to choose one of these four objects.

After this HIT has expired we will look at the choices of the other MTurkers
who also participated in this HIT using exactly the same screen as you.

Your choice will be compared with that of these MTurkers.
(as this HIT is restricted to the USA, the other MTurk participants will also be from the USA).

If the object you chose is **the same as the object chosen most frequently** by the other MTurkers,
you will receive **a bonus of $2.00** for this HIT (on top of your participation fee of $0.50).
Otherwise, you only receive your participation fee of $0.50.

#3 Decision screen

Please choose one of the objects below

Remember: in order to earn the bonus you need to choose the same object as most other MTurkers who participate in this HIT.
The other MTurkers will complete the task using exactly the same screen as yourself.

Main experiment – KW, SMILE-80px and SMILE-95px

#1 Welcome screen [all treatments]

Welcome

Thank you for participating in our HIT.
Including the time for reading these instructions, the HIT will take about 7 minutes to complete.

During the HIT, please do not close this window or leave the HIT's web pages in any other way.
If you do close your browser or leave the HIT, you will not be able to re-enter and we will not be able to pay you!

On top of your guaranteed participation fee of $0.50, you can earn a bonus of $2.00 dependent on your responses and the responses of other participants in this HIT.

You will receive a code to collect your payment via MTurk upon completion.

#2 Instructions 1/3 [all treatments]

Instructions 1/3

This HIT will ask how **socially appropriate** certain behavior is.
By socially appropriate, we mean behavior that you think most people would agree is the "correct" thing to do.

Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then other people might be angry at them.

#3 Instructions 2/3 [all treatments]

Instructions 2/3

You will receive a description of a situation in which a person must decide how to act.
This person is another MTurker from the USA taking part in a HIT. We refer to this person as "Person A".
You will be given a description of various possible actions Person A can choose to take.

After you receive the description of the situation, you will be asked to **evaluate each of the various possible actions** Person A can choose to take.

You must indicate, for each of the possible actions, whether taking that action would be "**socially appropriate**" or "**socially inappropriate**".
By socially appropriate, we mean behavior that you think most other MTurkers from the USA would agree is the "correct" thing to do.

In each of your responses, we would like you to answer as truthfully as possible,
based on your opinions of what constitutes socially appropriate or socially inappropriate behaviour.

#3a Instructions 3/3 [KW]

Instructions 3/3

You will next be given the description of the situation where Person A has to choose how to act.
In that situation, Person A can take 11 possible actions.
After you read the description of the situation, you must consider each possible action and evaluate how socially appropriate that action is.

You will evaluate each action using the following 4-point scale:

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

You will make evaluations for 11 actions in total.
Your **bonus payment** depends on how you evaluate these actions.

This works as follows:

After this HIT has expired we will look at the evaluations of the other MTurker who also participated in this HIT and evaluated the possible actions of Person A using exactly the same scale shown above.

The computer will then randomly select one of the 11 actions.
Your evaluation of this action will be compared with that of the other MTurkers
(as this HIT is restricted to the USA, the other MTurk participants will also be from the USA).

If your evaluation is the **same as the evaluation made most frequently** by the other MTurkers,
you will receive **a bonus of $2.00** for this HIT (on top of your participation fee of $0.50).
Otherwise, you only receive your participation fee of $0.50.

#3b Instructions 3/3 [SMILE-80px]

Instructions 3/3

You will next be given the description of the situation where Person A has to choose how to act.
In that situation, Person A can take 11 possible actions.
After you read the description of the situation, you must consider each possible action and evaluate how socially appropriate that action is.

You will evaluate each action using the following 4-point scale:

Very Socially Inappropriate

Somewhat Socially Inappropriate

Somewhat Socially Appropriate

Very Socially Appropriate

You will make evaluations for 11 actions in total.
Your **bonus payment** depends on how you evaluate these actions.

This works as follows:

After this HIT has expired we will look at the evaluations of the other MTurker who also participated in this HIT and evaluated the possible actions of Person A using exactly the same scale shown above.

The computer will then randomly select one of the 11 actions.
Your evaluation of this action will be compared with that of the other MTurkers
(as this HIT is restricted to the USA, the other MTurk participants will also be from the USA).

If your evaluation is the **same as the evaluation made most frequently** by the other MTurkers,
you will receive **a bonus of $2.00** for this HIT (on top of your participation fee of $0.50).
Otherwise, you only receive your participation fee of $0.50.

#3b Instructions 3/3 [SMILE-95px]

Instructions 3/3

You will next be given the description of the situation where Person A has to choose how to act.
In that situation, Person A can take 11 possible actions.
After you read the description of the situation, you must consider each possible action and evaluate how socially appropriate that action is.

You will evaluate each action using the following 4-point scale:

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

You will make evaluations for 11 actions in total.
Your **bonus payment** depends on how you evaluate these actions.

This works as follows:

After this HIT has expired we will look at the evaluations of the other MTurker who also participated in this HIT and evaluated the possible actions of Person A using exactly the same scale shown above.

The computer will then randomly select one of the 11 actions.
Your evaluation of this action will be compared with that of the other MTurkers
(as this HIT is restricted to the USA, the other MTurk participants will also be from the USA).

If your evaluation is the **same as the evaluation made most frequently** by the other MTurkers,
you will receive **a bonus of $2.00** for this HIT (on top of your participation fee of $0.50).
Otherwise, you only receive your participation fee of $0.50.

#4 Situation [all treatments]

The situation

This is the situation that you will evaluate.

Person A is taking part in one of our HITs where he/she can earn money.
Person A receives the following instructions for the HIT (in *italics*).

Person A's instructions

In this HIT, you are matched with another MTurker, whom we will call Person B.

You are endowed with $10 and Person B is endowed with $0.

Your task is to allocate the $10 dollars you are endowed with between you and Person B.

You can decide to allocate $10 any way you like in $1 increments. You can allocate all $10 to yourself and $0 to Person B; you can allocate some part of money to yourself and some to Person B, or you can allocate $0 to yourself and all $10 to Person B.

Your decision determines how much you and Person B will earn. If you decide to allocate $X to yourself and $(10-X) to Person B, you will earn $X and Person B will earn $(10-X).

After making your decision, Person B will be informed of the choice made.

#5 Action Intro [all treatments]

Evaluate Person A's possible actions

In the next screens you will find the 11 actions Person A can take to split the $10 between him/herself and Person B.

Please indicate for each action how socially appropriate you think the action is.

Remember: in order to earn the bonus you need to evaluate the actions in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the actions using exactly the same scale as yourself, as you see it on your screen.

#5a Decision Screen (GIVE 0) [KW]

Evaluate Person A's possible actions (1/11)

Person A allocates $10 to him/herself and $0 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

#5b Decision Screen (GIVE 0) [SMILE-80px]

Evaluate Person A's possible actions (1/11)

Person A allocates $10 to him/herself and $0 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

#5c Decision Screen (GIVE 0) [SMILE-95px]

Evaluate Person A's possible actions (1/11)

Person A allocates $10 to him/herself and $0 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

#6a Decision Screen (GIVE 1) [KW]

Evaluate Person A's possible actions (2/11)

Person A allocates $9 to him/herself and $1 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

#6b Decision Screen (GIVE 1) [SMILE-80px]

Evaluate Person A's possible actions (2/11)

Person A allocates $9 to him/herself and $1 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

#6c Decision Screen (GIVE 1) [SMILE-95px]

Evaluate Person A's possible actions (2/11)

Person A allocates $9 to him/herself and $1 to Person B.

Please indicate how socially appropriate you think this action is.

Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate

Remember: in order to earn the bonus you need to evaluate the action in the same way as most other MTurkers who participate in this HIT.
The other MTurkers will evaluate the action using exactly the same scale as yourself, as you see it on your screen.

(The other 9 decision screens are similar, only varying the amount allocated to person B)