WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Kai Barron
Tuomas Nurminen

Nudging cooperation

Discussion Paper

SP II 2018–305

December 2018

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Copyright remains with the authors.

Discussion papers of the WZB serve to disseminate the research results of work in progress prior to publication to encourage the exchange of ideas and academic debate. Inclusion of a paper in the discussion paper series does not constitute publication and should not limit publication in any other venue. The discussion papers published by the WZB represent the views of the respective author(s) and not of the institute as a whole.

Affiliation of the authors:

Kai Barron, WZB
(kai.barron@wzb.eu)

Tuomas Nurminen, Hanken School of Economics, Helsinki
(tuomas.nurminen@hanken.fi)

Abstract

Nudging cooperation[*]

This paper experimentally studies two simple interventions aimed at increasing public goods provision in settings in which accurate feedback about contributions is not available. The first intervention aims to exploit lying aversion by requiring subjects to send a non-verifiable ex post announcement about their contribution. The second intervention aims to nudge participants to higher contribution levels by simply labeling contributions of 16 or above as being ‘good’. We find that the ex post announcement mechanism does not have a significant effect on the cooperation rate. However, the nudge leads to a striking increase in the cooperation rate. We provide suggestive evidence that the nudge we use provides subjects with a focal point, helping conditional cooperators to coordinate their contributions. Moreover, despite the lack of monetary incentives to lie, we find that a significant minority of subjects inflate their announcements.

Keywords: cooperation, nudge, focal point, public good, lying, experiment

JEL classifications: C91, C72, H41, Z13

[*] The authors would like to thank Puja Bhattacharya, Steffen Huck, Erin Krupka and Topi Miettinen, as well as conference participants at the 15th Workshop on Social Economy for Young Economists in University of Bologna for helpful comments and suggestions. Furthermore, we thank Markus Konrad for help with programming in o-Tree, and Nina Bonge for excellent support in running the experiments. Tuomas Nurminen gratefully acknowledges financial support from the OP Group Research Foundation, the Society of Swedish Literature in Finland and the Yrjö Jahnsson Foundation.

1 Introduction

Contributing zero in the linear public goods game is a dominant strategy, yet in many circumstances we observe positive levels of cooperation (Ledyard, 1997). This has generated a large literature that has sought to understand the mechanisms that sustain and erode cooperation (see Chaudhuri (2011) for a review). However, many of the mechanisms that might normally sustain cooperation are ruled out in contexts with low observability—i.e. where it is costly or impossible to observe individual contribution levels.[1]

In this paper, we study how two fairly weak, but easily implementable, manipulations of the choice environment affect cooperation among strangers in a one-shot public goods game. In the resulting two-by-two experimental design, we consider two possible interventions suitable for encouraging cooperation in social dilemma contexts with low observability.

In the first treatment dimension, NUDGE, we ask whether providing a simple cooperation nudge can induce increased contributions.[2] More specifically, in the instructions for the NUDGE treatment, we provided subjects with a salient focal point that unambiguously divided the decision space in two—the socially desirable contribution levels above the focal point, and the socially undesirable below. The motivation for the nudge draws on insights from two distinct literatures. Firstly, moral suasion and recommendations have been found to increase contributions in certain experimental public goods games (Croson and Marks, 2001; Dal Bó and Dal Bó, 2014). Secondly, provision of a focal point may help to resolve the coordination problem that arises when a public goods game is played by individuals motivated by reciprocity (Schelling, 1960; Rabin, 1993; Mehta et al., 1994).[3]

In the second treatment dimension, ANNOUNCE, we ask whether simply requiring individuals to announce how much they contributed ex post could increase contribution levels. The motivation

[1]Such mechanisms include monetary and non-monetary punishment (Fehr and Gächter, 2000; Masclet et al., 2003; Cinyabuguma et al., 2005), reputational and social image concerns (Andreoni and Petrie, 2004; Rege and Telle, 2004), and endogenous and exogenous assortative group formation (Page et al., 2005; Gunnthorsdottir et al., 2007), among others.

[2]"Nudging" refers to the act of subtly designing the non-pecuniary elements of the choice environment so as to increase the likelihood of desired behavior. The seminal reference is Thaler and Sunstein (2008).

[3]More specifically, this assertion is a consequence of the following logical steps: (i) that a public goods game may constitute a coordination game when considering social utiliy (Rege and Telle, 2004), which is supported by substantial evidence documenting the existence of *conditional cooperators* (see, e.g., Fischbacher et al. (2001)), and (ii) in coordination games the provision of a focal point can help individuals to coordinate (Schelling, 1960; Mehta et al., 1994). Combining these two insights suggests that the provision of a focal point could help subjects to coordinate in a public goods game setting.

for this stems from the recent literature on lying aversion[4], as well as the literature on shame and guilt.[5] In particular, we posit that the announcement could increase contribution levels if (a) subjects are averse to lying, and (b) if they suffer a psychological cost (e.g. guilt) when (truthfully) announcing a low contribution level.[6] The first condition implies a cost for free-riding and lying about it, while the second condition implies a cost for free-riding and reporting it truthfully. Taken together, if an individual anticipates these costs, one option available to her is to increase her contribution level to avoid both: (i) lying, and (ii) the psychological cost of announcing a low contribution.[7]

Strikingly, we find strong support for the hypothesis that a simple nudge increases contributions. Subjects contribute over 40% more on average when provided with a nudge towards cooperating. Furthermore, in both NUDGE treatments, the median subject contributes exactly the value of the focal point (80% of her endowment). This is in a stark contrast to a median below 40% of the endowment in the two treatments without the focal point. We show that this upward shift in contributions is accompanied by an upward shift in beliefs about others' contribution levels. This is consistent with the explanation that the focal point helps participants to coordinate.

Interestingly, we find that ex post announcements have no influence on contribution levels. However, surprisingly, we find that a non-negligible proportion (13%) of all individuals lie about their contribution level, even though there are no monetary, strategic, or social image incentives to do so. Furthermore, of those who chose the Nash equilibrium strategy of contributing zero to the public good, 27% then lied about their contribution level when making their anonymous announcement. This is striking in light of the large literature that finds unambiguously that a large proportion of individuals prefer to tell the truth even when it is harmful to them, and/or others (Erat and Gneezy, 2012; Abeler et al., 2014; Gneezy et al., 2018; Abeler et al., 2018). To our knowledge, we are the first to document that a significant minority of subjects are willing to lie although there are no direct or indirect (strategic or monetary) incentives to do so. One potential explanation is

[4]See, e.g, Gneezy (2005); Erat and Gneezy (2012); Fischbacher and Föllmi-Heusi (2013); Gibson et al. (2013); Serra-Garcia et al. (2013); Abeler et al. (2014); Gneezy et al. (2018).

[5]See, e.g., Tangney et al. (1996); Smith et al. (2002); Battigalli and Dufwenberg (2007); Ellingsen et al. (2010); Battigalli et al. (2013). We use the term 'guilt' in a colloquial sense, referring to the negative affect one experiences after failing to meet one's *personal* moral standards. However, guilt in the sense of Battigalli and Dufwenberg (2007) also fits our setting.

[6]The psychological cost we have in mind is akin to the cost one would repeatedly experience if one refused to give money to a beggar and was afterwards reminded of the refusal on a daily basis.

[7]Importantly, in order to focus solely on the role played by lying aversion and shame/guilt concerns, we implemented a fairly weak version of the ex post announcement intervention. In particular, we ruled out strategic concerns, social image concerns and reputational concerns by using announcements that are non-verifiable and anonymous in a one-shot public goods game.

that for these subjects the psychological cost of announcing their selfish choice is higher than the cost of telling an anonymous lie.

The rest of the paper is organized as follows. Section 2 describes the experimental design, while the theoretical predictions are derived in Section 3. Section 4 presents our results, followed by a discussion in Section 5. Section 6 concludes.

2 Experimental design

2.1 The public goods game

Our experiment is embedded in a standard linear public goods game. Public goods games reflect the central characteristics of many economically important situations in which individuals face a tension between advancing their private interests and cooperating for the benefit of the group. We study possible channels for inducing cooperation in the subset of social dilemma situations where it is difficult to observe individual contribution levels since these are precisely the situations where it is most challenging to induce cooperation. Furthermore, the use of a public goods game allows us to study the underlying psychological motivations for cooperation relevant for a wide range of important social interactions.

In the experiment, subjects are randomly assigned to groups of four. Every subject is endowed with 20 experimental points and has to decide how many of these to contribute to a group project and how many to keep for herself. Every point kept increases individual earnings by one point. The sum of all points contributed to the group project is multiplied by 1.4 and divided equally among the four group members. Thus, subject i's monetary payoff from the game is given by:

$$\pi_i^m = 20 - g_i + 0.35 \sum_{j=1}^{4} g_j, \tag{1}$$

where g_j denotes the contribution to the group project by group member j.

The marginal per capita return (MPCR) of a contribution to the group is 0.35. Since it is less than unity, the dominant strategy for a selfish subject is to contribute nothing. Thus free-riding by all group members is the unique Nash equilibrium of the game in terms of material gain.[8] In the

[8]The Nash equilibrium payoff is 20 points for all group members. By contrast, since $4 \cdot 0.35 > 1$, the socially efficient outcome is that every group member contributes their full endowment, which results in individual earnings of 28 points. An individual subject's payoff is maximized if she contributes zero points to the group project but the rest of her group contributes full, in which case the free-rider earns 41 points.

treatment conditions below, we study whether tweaking the choice environment can induce higher cooperation levels by magnifying various intrinsic motives.

2.2 Treatments

We implemented four treatments: BASELINE, ANNOUNCE, NUDGE, and NUDGE + ANNOUNCE.[9] In each treatment, subjects played the linear public goods game described above once. After the public goods game, subjects completed a short survey that gathered demographic information, as well as information about their social preferences. Feedback about earnings was only provided at the very end of the experiment. In all treatments, subjects were required to answer a series of control questions to ensure comprehension before making their contribution decisions.

Table 1: Treatment Conditions

	No announcement	Ex post announcement
Neutral instructions	**BASELINE**	**ANNOUNCE**
Nudging instructions	**NUDGE**	**NUDGE + ANNOUNCE**

In the BASELINE treatment, instructions for the game were written using neutral terms. The NUDGE treatment was exactly the same as BASELINE with the exception of the following adjustments aimed at inducing a valenced focal point. At the end of the instructions for BASELINE, subjects were additionally told that if everyone contributes to the group project it would be beneficial for the group. A contribution of "16 or greater" would be labelled as a "good" contribution, and a contribution of "15 or below" would be labelled as a "bad" contribution. Finally, "good" contributions were written in green, while "bad" contributions were written in red. (Please refer to the instructions in Appendix C for further details.) This framing serves two purposes. First, it highlights the moral dimension of the contribution decision. Second, it introduces a salient focal contribution level of 16 that unambiguously divides the decision space into socially desirable and socially undesirable contributions.

[9] See Appendix C for instructions for treatment NUDGE + ANNOUNCE. The full set of instructions is available from the authors.

In the ANNOUNCE treatment, subjects were told that they would be asked to announce how much they contributed after making their contribution decisions. This announcement was anonymous and non-verifiable by other subjects. Individual announcements were then revealed to the other group members on the computer screen, with anonymous player labels. Subjects were therefore free to announce any integer in the interval $[0, 20]$. Importantly, the announcement procedure was described to the subjects in the instructions *before* the contribution decisions were made, so subjects could anticipate having to make the announcement.

The NUDGE + ANNOUNCE treatment was identical to the NUDGE treatment, with the addition of an announcement stage that was the same as the one in the ANNOUNCE treatment. The objective of this treatment was to test for the presence of an interaction effect between the two mechanisms targeted in the NUDGE and ANNOUNCE treatments.

2.3 Belief elicitation

It has been well documented that people tend to be *conditional cooperators* in public goods games, with their contribution level depending on their belief about others' contribution levels. It is therefore important to measure how subjects' beliefs shift across the different treatment conditions. In particular, the argument for providing a focal point to shift contribution levels relies directly on a shift in subjects' beliefs regarding others' contribution levels (thereby facilitating coordination at the focal point by *conditional cooperators*).

In order to better understand the mechanisms driving changes in contribution levels, we therefore elicited subjects' first order beliefs regarding the contribution choices of their group members. More specifically, in all treatments after the contribution decision, we first elicited subjects' beliefs about the *average* contribution of the other three group members. In addition, in the ANNOUNCE and the NUDGE + ANNOUNCE treatments we also elicited subjects' beliefs about the *individual* contributions of each of their group members. This was done after the contribution announcements were made and anonymously revealed to the group. This second belief elicitation allows us to analyse how beliefs react to the announcements.

We incentivized both belief elicitation tasks using the commonly used quadratic scoring rule (QSR) reward function.[10] Specifically, in the first belief elicitation the reward for subject i with a

[10] We chose to provide incentives for reporting accurate beliefs despite an earlier finding that reported beliefs in a public goods game are only marginally more accurate when the elicitation is incentivized (Gächter and Renner, 2010). Since the contribution stage precedes both belief elicitation tasks, and since the elicitation tasks were unexpected at the time of making the contribution decision, providing incentives for accurate beliefs should not affect contribution decisions.

stated belief $\bar{b}_i \in [0, 20]$ was

$$\pi_i^{B1} = 8 - 8\left(\frac{\bar{b}_i - \bar{g}_{-i}}{20}\right)^2,$$

where $\bar{g}_{-i}$ denotes the average contribution by the other three group members.

In the second belief elicitation, the reward for subject i with a stated belief b_i^j about group member j's contribution was

$$\pi_{i,j}^{B2} = 8 - 8\left(\frac{b_i^j - g_j}{20}\right)^2.$$

However, in the second belief elicitation subjects were informed that only one of the three stated beliefs would become payoff relevant. The payoff relevant belief was chosen randomly.

After the belief elicitation stage, the subjects were asked to complete a short survey on other-regarding preferences and demographics. Finally, each subject's earnings from the public goods game and the belief elicitation(s) were summed and converted to euros using an exchange rate of 4 points = 1 EUR.

2.4 Procedures

We conducted fourteen sessions with 24 subjects in the WZB-TU experimental laboratory in Berlin. In each session, we implemented one treatment condition. A total of 336 subjects, predominantly students in universities in Berlin, participated in the experiment.[11] Participants were solicited through an online database using ORSEE (Greiner, 2015), and the experiment was run using the experiment software oTree (Chen et al., 2016). In each session, subjects participated in two experiments. First, they participated in an investment game that is completely unrelated to the current paper, and second they played the one-shot public goods game studied here.[12] Sessions lasted up to 90 min, and participants earned, on average, 6 EUR for the public goods game, 1.8 EUR for the first belief elicitation, and 1.8 EUR for the second belief elicitation. In addition, they received a 5 EUR showup fee and their earnings from the investment game.

[11] There were 96 participants in each of the two ANNOUNCE treatments, and 72 participants in the other two treatments with no announcement stage. The rationale for having an additional session for each of the announcement treatments was to facilitate studying lying behavior, which could only be observed in these two ANNOUNCE treatments, and has a relatively low baseline level.

[12] The treatments of the two experiments were chosen independently of one another, and it was explicity and clearly stated to subjects that the two games were completely independent of one another. The first game did not study social preferences nor social norms, and feedback on earnings from it was only disclosed at the end of the whole session.

3 Predictions

In this section we outline our main hypotheses for each of the treatment conditions. Appendix A provides a more detailed discussion through means of a simple model.

3.1 Nudging

To understand how how our treatment manipulations translate into contribution decisions, we consider a simple model of individual decision making which incorporates the idea that an individual might experience disutility when her contribution level diverges from a benchmark personal reference contribution level. More specifically, we assume, firstly, that an individual cares about her monetary earnings, π_i, which is given by equation 1.[13] Secondly, she may suffer a disutility $\lambda_i(g_i - N_i)^2$, if her contribution, g_i, differs from her reference contribution, N_i. The parameter $\lambda_i \geq 0$ represents the weight the individual puts on this intrinsic component of her utility. The personal reference contribution level may reflect either the individual's view of what is the morally appropriate contribution level, or of what her peers contribute on average.[14] Thus, we assume that an individual maximizes the expectation (with respect to the contributions by her group members) of the utility function:

$$U_i(g_i, g_{-i}; N_i) = \pi_i^m(g_i, g_{-i}) - \lambda_i(g_i - N_i)^2, \tag{2}$$

in which (g_i, g_{-i}) is the vector of contributions by i's group.

In the instructions for the NUDGE treatment, contributions of "16 or greater" were labelled as "good" and contributions of "15 or below" as "bad". We expect this nudge to essentially have two effects. Firstly, appealing to the normative nature of the contribution decision may affect subjects' preferences, independent of what they think others will do. Specifically, we posit that normative appeals increase the contribution level that subjects consider appropriate. Secondly, the nudge likely affects subjects' beliefs about others' contributions. Based on previous literature, we expect the beliefs to shift upwards, on average, and at the same time largely cluster at the natural focal point of 16. Therefore, in terms of our model, we expect the NUDGE treatment to increase the expected value of N_i. This leads to an increase in the average contribution. In Appendix A

[13] For simplicity, we assume that the individual is risk neutral with respect to her own monetary earnings.

[14] In the former case N_i is fixed and known to the individual ex ante. Hence, the individual can be seen as following a form of deontological ethics. In the latter case, N_i is the average contribution by the individual's group members, and is thus only known to the individual ex post, which means she has to act on the basis of her expectation of N_i. An individual of this type is thus motivated by reciprocal concerns (Rabin, 1993).

we derive the result formally. The following summarizes our hypotheses regarding the treatment effect on contribution levels in the NUDGE treatment.

> **Hypothesis 1.** *(i) The average contribution is higher in the* NUDGE *treatment relative to the* BASELINE *treatment. (ii) The fraction of contributions above sixteen is higher and the fraction of intermediate level contributions (i.e. between 1 and 15) is lower in the* NUDGE *treatment relative to the* BASELINE *treatment.*

3.2 Ex post announcements

In the ANNOUNCE treatment, we add anonymous non-verifiable ex post announcements to test whether an aversion to lying can help to sustain cooperation. The basic intuition of this treatment is that when we introduce the ex post announcements, individuals may face two additional psychological motives relative to the BASELINE treatment, namely (i) a cost of lying, and (ii) a psychological cost (e.g. shame or guilt) to announcing a selfish low contribution level. The anticipation of these two costs may influence behavior. If the two psychological costs are sufficiently high to dominate the monetary benefits of making a low contribution, the individual may increase her contribution level. Alternatively, she may leave her contribution level unaffected, and pay one of the two psychological costs. We summarize our hypothesis as follows:

> **Hypothesis 2.** *The average contribution in the* ANNOUNCE *treatment is higher than the average contribution in the* BASELINE *treatment.*

3.3 Interaction effects

Lastly, we test for an interaction effect between the two mechanisms under consideration. We posit that there are two mutually exclusive possibilities. On the one hand, if the NUDGE treatment is effective at increasing the intrinsic incentives to choose a contribution level above 16, then by the same psychological rationale, it may also increase the psychological cost of (truthfully) announcing a contribution level below 16. This implies that the effect of introducing the announcement mechanism is more pronounced when combined with the moral nudge than without it. This logic leads to Hypothesis 3a.

> **Hypothesis 3a.** *The increase in average contributions in the* NUDGE + ANNOUNCE *treatment relative to the* NUDGE *treatment is larger than the increase in average contributions in the* ANNOUNCE *treatment relative to the* BASELINE *treatment.*

On the other hand, if the NUDGE treatment is very effective, there might not be much scope for improvement anymore in the NUDGE+ANNOUNCE treatment. This implies that the effect of introducing the announcement mechanism is more pronounced without the moral nudge than with it, generating Hypothesis 3b.[15] We summarize this idea in the following hypothesis that is complementary to 3a (i.e. Hypotheses 3a and 3b are mutually exclusive).

Hypothesis 3b. *The increase in average contributions in the* ANNOUNCE *treatment relative to the* BASELINE *treatment is larger than the increase in average contributions in the* NUDGE + ANNOUNCE *treatment relative to the* NUDGE *treatment.*

Therefore, in the scenario where the NUDGE treatment alone has a small effect relative to the BASELINE, hypothesis 3a is the appropriate one. However, in the scenario where the NUDGE treatment alone has a large effect relative to the BASELINE, hypothesis 3b is the appropriate one. Essentially, if the nudge alone succeeds in achieving high contribution levels, there remains no scope for a strong interaction effect.

4 Results

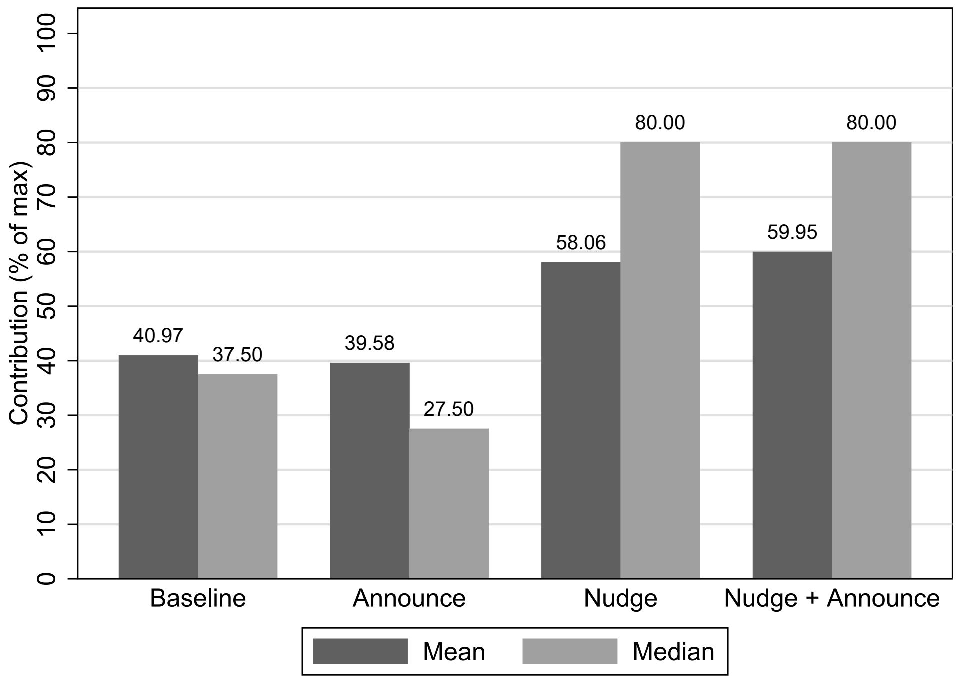
We start by examining how contribution choices are influenced by the treatment condition. Thereafter, we provide insight into the mechanisms driving the observed pattern of contributions by focusing on the roles played by beliefs and lying aversion in shaping behavior. Table 2 provides a descriptive overview of the data. Figure 1 displays the mean and median contribution levels in each treatment, denoted as a percentage of the maximum possible contribution. Consistent with the prior literature, in the BASELINE treatment we observe the majority of subjects making positive contributions, with an average contribution level of 41%. However, it is immediately apparent from the figure that nudging cooperation has a large influence on contribution levels, increasing the average contribution by over 40% to a contribution level of 58% in the NUDGE treatment. Furthermore, the median contribution doubles from 40% to 80% to lie exactly at the focal contribution level of 16. A one-sided Wilcoxon rank-sum test (WRS) confirms that the increase in contributions in the NUDGE treatment relative to the BASELINE treatment is highly statistically significant ($p = 0.004$). This suggests that a substantial proportion of subjects are increasing their contribution level to coordinate at (or above) the focal point.

In contrast, the ex post announcements appear not to have shifted average contribution levels. There is no increase in mean contributions in the ANNOUNCE treatment relative to the BASELINE

[15] Again, we refer to interested reader to Appendix A for further details.

treatment (WRS, one-sided, $p = 0.63$), nor in the NUDGE + ANNOUNCE treatment relative to the NUDGE treatment (WRS, one-sided, $p = 0.48$). One reason for this is that since the announcements were anonymous, the intrinsic guilt/shame costs of announcing a low contribution level were perhaps not large enough to outweigh the material costs of choosing a higher contribution level.

Figure 1: Contribution levels across treatments (mean and median)



Having demonstrated the pattern of contributions generated by the treatments, the discussion below sheds some light on the channels driving this pattern of behavior.

In Section 4.1 we show that the increase in contribution levels generated by providing a nudge is associated with an upward shift in beliefs of similar magnitude. This evidence is consistent with the explanation that many participants are *conditional cooperators*, and the focal point simply provides them with an opportunity to coordinate.

Section 4.2 studies part (ii) of Hypothesis 1 by presenting evidence regarding the influence that the treatment has on the distribution of contributions. In particular, we show that, consistent with Hypothesis 1(ii), nudging cooperation operates predominantly by shifting individuals who would otherwise have contributed between 5% and 79% to increase their contribution level to above 80%. In particular, the proportion contributing nothing remains relatively stable.

Section 4.3 turns attention to studying lying behavior, seeking to explain why Hypotheses 2 and 3 were rejected. In particular, we show that among individuals who contribute nothing, more than 25% choose to lie about their contribution. This is striking as these subjects have no monetary or strategic incentives to lie.

Table 2: General descriptive statistics

	N	Average contribution	Average ex ante belief	Percentage lying	Average lie among liars	Average ex post belief
BASELINE	72	41.0 (36.0)	48.6 (25.2)			
ANNOUNCE	96	39.6 (38.8)	47.8 (27.4)	12.5	57.5 (31.9)	41.6 (24.1)
NUDGE	72	58.0 (40.7)	63.8 (26.6)			
NUDGE + ANNOUNCE	96	60.0 (39.0)	67.7 (24.6)	13.5	50 (41.6)	61.4 (22.4)

Notes: (i) Standard deviations are shown in brackets.
(ii) Contributions, beliefs and lie magnitudes are expressed as a percentage of the maximum.

4.1 The role of beliefs in cooperating

Several models have been suggested for understanding contribution choices in the realm of social dilemmas. An important class of these models are motivated by the empirical observation that subjects appear to behave like *conditional cooperators*, making their contribution choices as an increasing function of their beliefs about others' contributions (Keser and Van Winden, 2000; Fischbacher et al., 2001; Frey and Meier, 2004). This class includes models focusing on distributional concerns over outcomes (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and models of intentions based reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). For these models, subjects' beliefs about others' actions are an important determinant of contributions in the public goods game. The models would thus predict that providing a focal point would allow subjects to coordinate on it by shifting both their beliefs about others' contributions, and their own contributions to the focal point. In this section, we therefore study the role played by beliefs about others' contribution choices, in order to assess whether behavior is consistent with this class of models and to gain greater insight into the underlying mechanism driving the higher contributions in the NUDGE treatments.

Table 3 reports the regression estimates showing the influence of the treatments on contribution levels and ex ante beliefs (i.e. prior to receiving the announcements from group members).[16] In particular, column (1a) shows that providing the nudge increased average contributions by 17.1 percentage points. Interestingly, column (1b) shows a similar upward shift in subjects beliefs

[16]In Table 3, the treatment dummy variables are defined as follows. $T_N = 1$ for individuals in the NUDGE or NUDGE + ANNOUNCE treatments. Similarly, $T_A = 1$ for individuals in the ANNOUNCE or NUDGE + ANNOUNCE treatments. $T_N * T_A = 1$ for individuals in the NUDGE + ANNOUNCE treatment.

about the average contribution level of their group members of 15.1 percentage points. Columns (1a) and (1b) therefore provide suggestive evidence that the treatment effect is mediated by shifting beliefs, which then induces higher contributions by the *conditional cooperators*. This is further supported by the results in column (2a), which show that a one percentage point increase in an individual's ex ante belief is associated with a one percentage point increase in her contribution choice. Furthermore, column (2b) shows that after controlling for beliefs, the nudge has no further effect on contributions.

This evidence is consistent with the idea that the treatment works by allowing individuals to coordinate on a commonly known focal point. It does this by shifting beliefs about what others will do, thereby leading to a shift in one's own contribution choices.

Table 3: Treatment Effect

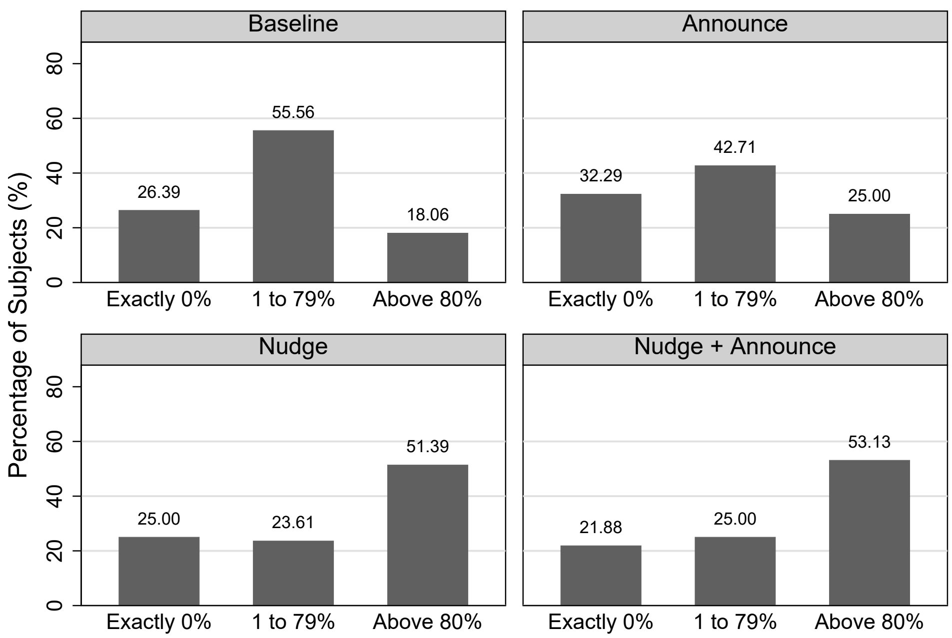
	Contributions (1a)	Ex Ante Beliefs (1b)	Contributions (2a)	Contributions (2b)
<u>Treatments</u>				
Nudge ($T_N = 1$)	17.1***	15.1***		1.6
	(6.45)	(4.35)		(5.11)
Announcement ($T_A = 1$)	-1.4	-0.9		-0.5
	(5.84)	(4.10)		(4.35)
Nudge * Announcement ($T_N * T_A = 1$)	3.3	4.8		-1.6
	(8.57)	(5.76)		(6.28)
Ex ante belief			1.0***	1.0***
			(0.05)	(0.06)
Constant	41.0***	48.6***	-8.9**	-8.7**
	(4.28)	(2.99)	(3.16)	(3.97)
Observations	336	336	336	336
Adjusted R^2	0.049	0.102	0.500	0.496

Notes: (i) Robust standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)
(ii) The regression includes a binary dummy variable for the nudge treatment, a binary dummy variable for the announce treatment, and a dummy for the interaction effect (as opposed to a dummy variable for each treatment).

4.2 Heterogeneity in contribution distributions

In Hypothesis 1(ii) we posited that the nudge would impact predominantly individuals who receive some intrinsic reward from contributing, and therefore would have been likely to contribute more than zero in the BASELINE treatment (see also Appendix A). This implies that the treatment effect would operate primarily by shifting individuals who were contributing between 1 and 15 (i.e. 5% to 75%) to a contribution level of at least 16 (i.e. 80%). Figure 2 shows that this hypothesis is supported in the data.[17] In particular, the proportion of individuals contributing zero remains relatively stable across the four treatment groups, averaging 23.1% in the two NUDGE treatments, and 29.8% in the two NON-NUDGE treatments. This finding is consistent with our explanation that the NUDGE treatment has no impact on individuals who are not motivated by other-regarding concerns.[18]

Figure 2: Distribution of contribution levels



On the contrary, the proportion contributing above the focal point (i.e. at least 80%) more than

[17]In Figure 5 in Appendix B we provide the full distribution of contributions in each treatment. A Kruskal-Wallis test confirms that the (population) contribution distributions are significantly different across treatments ($p < 0.001$).

[18]The proportion of selfish money-maximizers in our experiment is in line with previous findings. For example, Fischbacher et al. (2001) observe that around 30% of their subjects were (unconditional) free riders. Similarly, Andreoni and Miller (2002) find that about 25% of their subjects were selfish money-maximizers.

doubles as a consequence of the nudge, averaging 52.4% in the two treatments with the nudge and 22.0% in the two treatments without it. This implies that the proportion contributing between 5% and 75% is significantly lower in the two treatments with the nudge than in the two treatments without it.[19]

Rege and Telle (2004) consider a similar analysis of the heterogeneity of their treatment effect. Like us, they observe a fairly stable proportion of non-contributors, and that their social approval treatment effect operates predominantly by shifting intermediate contributors to high contribution levels. While their social approval treatment is very different from our NUDGE treatment, the parallels between the results in terms of how treatment influences individuals at different parts of the distribution are interesting. One explanation is that the intermediate contributors are *conditional cooperators* who receive intrinsic rewards from cooperating and who simply hold low beliefs about the contribution levels of others. In this case, both the social approval treatment of Rege and Telle (2004) and our NUDGE treatment shift beliefs about others' contribution levels, facilitating coordination.[20]

4.3 Lying for self-image

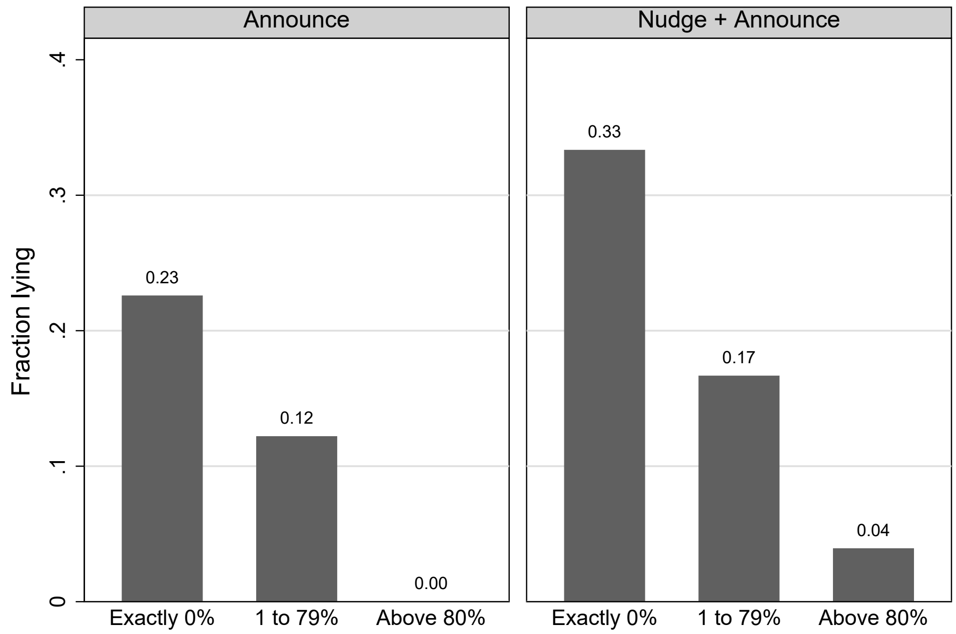
In our predictions section above, we discuss how individuals who have an intrinsic psychological cost of announcing a selfish, low contribution have two avenues available to them to avoid making this low announcement. They can either increase their contribution level in anticipation of the shame cost, thereby avoiding it. Or, they can make a low contribution and simply lie about it. Above, we showed that the ANNOUNCE treatment had no influence on contribution levels, suggesting that the shame cost of the non-contributors was not sufficiently high to affect their contribution levels. This is perhaps not extremely surprising given that we chose a relatively weak version of these announcements - i.e. anonymous and non-verifiable. However, given the subtlety of the intervention it is rather surprising that a non-negligible proportion of individuals are willing to lie to avoid making a low announcement. In particular, 13% of all subjects lie in their announcement.

[19]Three Fisher's exact tests confirm that (i) the proportion of subjects contributing nothing does not differ across treatments ($p = 0.43$), (ii) the proportion of subjects contributing between 1 to 15 points differs significantly across treatments ($p < 0.001$), and (iii) the proportion of subjects contributing above 16 points differs significantly across treatments ($p < 0.001$).

[20]In another interesting paper considering the heterogeneity in contribution distributions across treatments in a public goods game, Andreoni and Petrie (2004) introduce both social image concerns and categorical reporting. Essentially, subjects are asked to publicly announce whether they contributed 0-14 tokens, or 15-20 tokens. While the category reporting did not increase average contribution levels, it did shift the distribution of contribution choices away from 1-13, and towards 0 or 15. This is similar to the shift away from intermediate contribution choices observed in our experiment (although, we don't observe a shift towards 0).

However this includes all the individuals who chose high contribution levels and had no motive to lie. Therefore, it is more informative to examine Figure 3 which displays the proportion of individuals lying, split by their contribution choice.

Figure 3: Lying as a function of contribution choice



Focusing on the individuals who contributed zero, the rate of lying is 22.6% in ANNOUNCE and 33.3% in NUDGE + ANNOUNCE.[21] The rate of lying among free riders differs significantly from zero in both these treatments (two Binomial tests, $p < 0.001$). This is striking as these individuals have absolutely no monetary, strategic nor social image reason for lying. In view of the substantial literature documenting an aversion to lying, this suggests that for these individuals the guilt/shame costs of announcing their selfish choice of contributing nothing outweighed any aversion to lying they may have had. To the best of our knowledge, we are the first to document that a tension between intrinsic motives can induce lying, when the individuals' lying aversion is dominated by another intrinsic motive.[22]

[21] There are 31 individuals who choose to contribute zero in ANNOUNCE and 21 individuals who chose to contribute zero in NUDGE + ANNOUNCE.

[22] As described above, for subjects who would contribute nothing in the BASELINE treatment, we induce a conflict between their intrinsic inclination not to lie, and their intrinsic inclination not to announce their selfishness of contributing nothing.

Our finding that many people are willing to lie even if the announcements have no monetary payoff consequences poses a challenge to the literature on *partial honesty*, whereby individuals are assumed to (strictly) prefer to tell the truth if the communication does not affect their monetary payoff (Matsushima, 2008; Dutta and Sen, 2012; Kartik et al., 2014; Ortner, 2015).

4.3.1 Lying and gender

While we did not have a hypothesis about gender differences in lying behavior, we report here as a post hoc finding that of the 25 subjects who lied in our experiment 20 were men, 4 were women, and one subject chose not to answer the gender question. Our data was unbalanced with respect to gender, but the relative numbers point in the same direction: 17.4% of the men lied, while only 5.4% of the women lied. We can reject the null hypothesis that the decision to lie is independent of gender (χ^2 test, $p = 0.03$, and Fisher's exact test, $p = 0.02$). Thus, we find some evidence that men are more prone to lie in this context than women.

One potential simple explanation for this result could be that women contribute more, and therefore have lower (intrinsic) incentives to lie. This explanation would require no statements about a gender difference in the propensity to lie. However, even when we only look at those subjects who contributed below 16, and hence had an incentive to lie from the point of view of impression management, the difference in lying rates between men and women persists: 26.4% of the men who contributed below 16 lied, while only 7% of the women contributing below 16 lied (χ^2 test, $p = 0.02$, Fisher's exact test, $p = 0.01$).[23]

Our findings contribute to existing literature on gender differences in lying and truth-telling. This literature has obtained inconclusive results, with several studies finding that men are more prone to lie than women (Dreber and Johannesson, 2008; Friesen and Gangadharan, 2012; Abeler et al., 2014; Conrads et al., 2014), while others finding no significant differences in lying behavior between men and women (Childs, 2012, 2013; Gylfason et al., 2013). Muehlheusser et al. (2015) do not find gender differences when decisions are made individually, but when decisions are made in groups, women-only groups lie significantly less than men-only and mixed groups.

[23] Furthermore, since the majority of individuals who contributed below 16 in these two treatments, contributed zero, if we look at the lying rate for individuals who contributed exactly zero, the pattern is very similar. 16.7% of the 18 women who contributed zero, lied. In contrast 32.4% of the 34 men who contributed zero, lied about it. This difference is not significant due to the low number of observations.

4.4 Ex post beliefs

Recall that in treatments ANNOUNCE and NUDGE + ANNOUNCE, we elicited every subject's beliefs about the contribution of each of her group members *after* the announcements were made and revealed. This allows us to analyze how subjects update their beliefs after seeing the announcements. While belief updating is not the main focus of our study, it is interesting to see how subjects assess the credibility of others' announcements when they know that there do not exist monetary or strategic incentives to lie.

The average ex post belief is 41.6% in the ANNOUNCE treatment, and 61.4% in the NUDGE + ANNOUNCE treatment (a two-sided WRS, $p = 0.0003$).[24] The difference is hardly surprising, considering that both contributions and announcements were higher in NUDGE + ANNOUNCE than in ANNOUNCE.

More to the point, ex post beliefs are lower than ex ante beliefs in the ANNOUNCE treatment (a two-sided Wilcoxon signed-rank test (WSR), $p = 0.08$), as well as in the NUDGE + ANNOUNCE treatment (a two-sided WSR, $p = 0.03$). This suggests that ex post beliefs tend to be more accurate than ex ante beliefs. We test this claim formally by first constructing for each subject a measure of ex ante belief accuracy and three measures of ex post belief accuracy (one for each group member). We do this by subtracting the relevant belief from the actual contribution. Within each group, we then calculate the average ex ante belief accuracy and the average ex post belief accuracy. Average ex ante belief accuracy in the two treatments with announcements is -1.59 points, while average ex post belief accuracy is -0.34 points (a two-sided WSR, $p = 0.004$). While both ex ante and ex post beliefs tend to be, on average, too optimistic, ex post beliefs are generally much more accurate.[25]

This finding is interesting in that it shows that announcements do carry informational content. In general, subjects see the announcements as credible and update their beliefs accordingly, which results in largely realistic ex post judgments.

5 Discussion

This paper contributes to a rich literature studying how cooperation rates can be strongly influenced by contextual factors which trigger a particular psychological mechanism. In particular, making the contribution decision may bring several psychological motives into conflict (e.g. self-interest and fairness). The resolution of this internal conflict determines the final contribution choice. How-

[24] We respect the independence assumption by basing all our tests in this section on group averages.

[25] These observations hold even if we compare ex ante beliefs to ex post beliefs by treatment. In addition, both ex ante and ex post beliefs are equally accurate *across* the two treatments.

ever, external contextual factors can shift the individual's attention between the different motives, influencing the weight she places on each motive.[26] In the sections below, we discuss how this paper fits into the important literature studying how the choice environment influences cooperation rates in social dilemma games.

5.1 Framing, salience and experimenter demand effects

Several papers have studied the role played by framing and social norms in public goods provision (see, e.g., Andreoni (1995); Cookson (2000); Rege and Telle (2004); Dufwenberg et al. (2011)). These papers typically take one of two approaches: (i) comparing a 'give' versus a 'take' framing of the contribution choice, or (ii) comparing a neutral description of the game with a socially valenced 'community' description of the game in which terms such as 'cooperate' and 'free-ride' are emphasised.[27] Since all our treatments use a 'give' frame, our paper relates more closely to the 'community' framing literature. Within this literature, Rege and Telle (2004) is perhaps closest to the current paper, both in terms of the approach taken and their core research objectives.

While Rege and Telle (2004) differs from the current paper along several dimensions, which reduces the direct comparability of the contribution level choices, it is still informative to compare the treatment effects in the two papers.[28] In their setting, while the introduction of associative framing does shift the point estimate of the average contribution level upwards, this shift is not statistically significant.[29] In another study testing the influence of the 'community' description frame, Dufwenberg et al. (2011) find significantly *lower* ($p = 0.034$) contribution levels under the 'community' description frame, in comparison to the neutral frame. If one views our NUDGE treatments as simply varying the framing, and reinforcing an individual's internalised cooperation norm, then our results contrast with this evidence, since we find a strong upward shift in contributions in both

[26] Of course, the individual's internal psychological dispositions are also extremely important in determining the weight she places on different motives — contextual factors can only lead to marginal shifts in these weights for a given individual. The outcome of this conflict between motives manifests in the individual's choice behavior, which is usually taken as reflecting her preferences.

[27] In an interesting paper, Tsikas et al. (2018) frame a repeated linear public goods game as an income tax reporting game. The tax frame increased contributions, on average, by almost 40% compared to a neutral frame, but the effect practically vanished over time.

[28] The main differences in experimental design between the two experiments are the following. Unlike us, they employ a 'take' frame for all their treatments, and have subjects interacting in groups of 10, with a MPCR of 0.2.

[29] In a 2x2 factorial design, they study: (i) the role played by framing, and (ii) the influence of social image concerns. This provides two tests of the effect of introducing framing (i.e. with dimension (ii), social image concerns, turned on or off). First, in the absence of social image concerns the introduction of framing increased average contributions from 34.4% to 55.1%, but this is not a statistically significant increase (a Mann-Whitney U test, $p = 0.09$). Secondly, in the presence social image concerns, the introduction of framing increased contributions from 68.1% to 77.3%, which again was not a significant increase (a Mann-Whitney U test, $p = 0.25$).

of our NUDGE treatments.

However, as discussed above, our NUDGE frame differs in the sense that it provides a clear, simple, valenced focal point. Rather than providing an ambiguous prescription that 'contributing more is socially desirable', our treatment encodes the contribution space into two clear categories: contributions that are 'good' or socially desirable, and contributions that are 'bad' or socially undesirable. This serves to remove several sources of ambiguity. Firstly, it removes the internal psychological ambiguity that might arise when an individual introspectively tries to figure out where her personal line between normatively 'good' and 'bad' contributions lies (i.e. with respect to her intrinsic self-image reward from making a contribution to the social good). Secondly, it reduces ambiguity regarding the individual's beliefs about what others believe is a 'socially desirable' contribution, and consequently also the ambiguity regarding others' contribution choices.[30] Putting this another way, subjects are provided with a simple focal point on which to coordinate, and more than half do so (see the bottom two panels of Figure 2).

One potential concern that can be raised is that our observed treatment effect is in fact simply an artifact of the experimenter demand effect (see, e.g., Zizzo (2010) and De Quidt et al. (2018)). However, we posit that there are several factors that suggest that experimenter demand effects may not be a major concern in our experiment. Firstly, the evidence presented in Section 4.1 showed that the observed treatment effect in contribution levels moved in tandem with a shift in beliefs about others' contribution choices. This implies that any argument that our treatment effect is driven by experimenter demand effects must also contend that these experimenter demand effects are also shifting beliefs about others' contributions. This is not completely implausible but places a greater demand on the experimenter demand effect explanation of our results.[31] Secondly, the explicit objective of the NUDGE treatments is to generate a commonly shared encoding of the contribution space into two categories — contributions labeled as 'good' and contributions labeled as 'bad'. Therefore, the mechanisms generally used to explain the experimenter demand effect (e.g. social expectations regarding taking a specific action) are being explicitly targeted by this treatment.[32] Thirdly, since it is fairly clear that the socially desirable choice in most public goods

[30]We would conjecture that this ambiguity in the presence of conflicting motives opens the door for self-serving biases in reasoning to play a role, shifting contributions downwards.

[31]It requires, firstly, that some subjects increase their contribution levels because they feel cognitive or social pressure from the experimenter to do so. Secondly, it requires that the same set of subjects believe that all other subjects also increase their contribution choices due to feeling cognitive or social pressure to do so from the experimenter.

[32]A noteable exception to this is that one of the core mechanisms that fall under the umbrella term of experimenter demand effects is the idea that participants are responding to the expectations of an *authority figure* à la Milgram (1974). However, it is not obvious why this would particularly be an issue in the current experiment. And, moreover, why it would only operate in the NUDGE treatments.

games is a high contribution level, one can argue that any experimenter demand effect present in our experiment is also present in other public goods games experiments. Furthermore, it is perhaps not obvious a priori why our NUDGE treatments should exacerbate these experimenter demand effects more than our ANNOUNCE treatments, or the 'community' framing studied in related work.[33] Nevertheless, while these arguments provide reasons why experimenter demand effects are unlikely to be the predominant driver of our results, we cannot completely rule out the possibility that the subjects are responding to the perceived demands of the experimenter, and the experimental setting. So this remains a caveat to our results.

5.2 Communication in public goods games

The ANNOUNCE treatment variation in our experiment relates closely to the large literature on communication in social dilemma games (Isaac and Walker, 1988; Miettinen and Suetens, 2008; Bochet and Putterman, 2009; Balliet, 2010; Koessler et al., 2018).[34] The main finding of this literature is that *pre-play communication* facilitates cooperation (Bochet et al., 2006).[35] However, the objective of our ANNOUNCE treatment is somewhat different—we shut down many of the mechanisms that may be present in pre-play communication and focus only on the potential role played by self-image costs, such as *shame* and *lying aversion*, in helping to sustain cooperation in environments where it is difficult to observe others' contribution levels. In particular, by focusing on anonymous, non-verifiable *ex post communication*, we remove strategic concerns, detection concerns, social image concerns, and the possibility that communication helps subjects to understand the game.

There were several reasons for choosing this setting. Firstly, in many real-life situations in-

[33] Additionally, sometimes the experimenter demand effect is defined as arising when subjects' behavior is affected by what they think the experimenter desires. We note that while the instructions in the NUDGE treatments can be seen as a signal of the experimenter's desire for high contributions, contributing high also means that the experimenter is losing more money. This tension between the experimenter's desires should mitigate concerns that the treatment effect observed in the two NUDGE treatments is driven by such an experimenter demand effect.

[34] In an interesting study outside the domain of public goods games, Bhattacharya et al. (2017) explore the role played by communication in triggering similar psychological mechanisms to those we consider. In particular, in a principal-agent setting, the authors examine the influence of non-binding *ex ante* or *ex post* announcements made by the agent on her effort choices. While the setting is different, the paper also explores the role that communication can play in enhancing cooperation. In their setting, ex post communication is somewhat effective.

[35] Balliet (2010) succinctly summarizes some possible underlying mechanisms: "Several explanations for the effect of communication include a better understanding of the game, increasing expectations of cooperation, enhancing group identity, and generating norms of cooperation. [...] However, research has identified group identity and norms as the most likely explanations." In particular, communicating intentions or giving promises about future behavior have been found to be especially effective in enhancing cooperation (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008; Miettinen, 2013).

volving public goods provision, it is costly or impossible to view others' action choices.[36] Secondly, the public goods literature argues that many individuals behave like *conditional cooperators*, and in repeated game settings, with accurate feedback about others' contribution choices, the formation of beliefs about others' contribution levels is not extremely complicated. However, in low-information settings, where individuals need to rely on announcements by others to obtain information about contributions, it is less obvious how belief formation works and how these beliefs will affect contribution choices. What then becomes relevant is how players communicate in the game and how credible this communication is.[37] Thirdly, in order to better understand which underlying mechanisms make communication effective, it is necessary to introduce them one by one.[38]

In the domain of communication, our paper relates most closely to two recent papers by Bernd Irlenbusch and Janna Ter Meer who experimentally study the role played by non-verifiable ex post communication in a *repeated* public goods setting (Irlenbusch and Ter Meer, 2013, 2015). The authors thus take a first step in addressing the second point raised in the previous paragraph. Interestingly, they find high rates of lying in combination with low contribution choices, however subjects correctly anticipate that others are lying and accordingly form low beliefs about others' contribution choices. The main difference between their setting and ours is that since their game is repeated, subjects have a strong strategic incentive to try to convince others that they have been making high contributions. Therefore, the mechanisms explored in the two papers are very different. Our interest lies in the impact of the *existence* and *anticipation* of the ex post announcement mechanism rather than in the impact of the information communicated by the announcements and its influence on later rounds. Indeed, when strategic motives to lie are absent, we observe markedly lower lying rates than Irlenbusch and Ter Meer (2013, 2015). However, given the absence of material and social motives to lie, it is striking that we observe lying at all.

5.3 Recommendations and moral suasion

Our NUDGE treatment complements two distinct but closely related literatures that investigate the impact of two particular forms of pre-play communication from the part of the experimenter on

[36]One can think of examples ranging from working in a team, where it's hard to observe the effort of other team members, to fishing companies fishing from a common pool of fish, to countries exerting effort to fight climate change. In a repeated game setting, Neugebauer et al. (2009) and Ambrus and Greiner (2012) study public goods contribution choices when feedback between rounds is noisy or nonexistent.

[37]Using a combination of field and lab experiments in a charitable giving context, Kessler (2017) shows that merely announcing what one *would* contribute if one had an opportunity may induce others to contribute.

[38]It would be informative for future work to study the implications of gradually introducing: strategic concerns, detection concerns, and social image concerns in relation to ex post announcements.

contributions in public goods games. The first one examines the impact of recommended contributions (e.g. Croson and Marks (2001)), while the second one focuses on the effect of exogenous moral messages (e.g. Dal Bó and Dal Bó (2014)). Croson and Marks (2001) find weak or no effect of recommendations on contributions in a threshold public goods game, although these findings may be confounded by experimenter demand effects. Dal Bó and Dal Bó (2014) find that recommendations backed up by moral appeals have a positive impact on contributions over and above the impact of neutral recommendations or experimenter demand effects.

Our NUDGE treatment implicitly recommends a contribution of at least 16, and can thus be seen as a special type of recommendation. However, this recommendation is not morally neutral, as it unambiguously labels certain contributions as morally acceptable while others as morally unacceptable. Hence it constitutes a form of moral suasion, albeit not one that appeals to any universal maxim. As such, it is comparable to the two moral treatments in Dal Bó and Dal Bó (2014).

5.4 Social image and self-image concerns

The objective of our ANNOUNCE treatment was to magnify the influence of intrinsic motives which may play a role in influencing cooperation decisions in real world settings (e.g. lying aversion and personal guilt or shame from making a low announcement). The prior literature has studied the cooperation-enhancing properties of manipulating the image costs of low contributions, but this literature has focused predominantly on introducing social image costs. For example, in early work on this issue, Andreoni and Petrie (2004) and Rege and Telle (2004) study the implications of subjects publicly announcing their contribution choices. Both studies find that revealing subjects' identities linked to their choices increases contribution levels. Clearly, this evidence on the influence of social image concerns is interesting and important for a wide range of settings. However, the results of the current paper complement this work by exploring the possibility that more subtle self-image motives may also play a role in shifting contribution choices.

6 Conclusion

We report results from an experiment designed to investigate the impact of two relatively weak manipulations of the choice environment on cooperation in a low-information public goods game. Situations requiring voluntary cooperation under low observability abound in the real world. This creates a demand for soft methods that help to mitigate the tension between individually rational and socially optimal behavior in low-information environments where more direct methods are impossible or expensive to implement.

Our results show that providing the individuals with a moral nudge and an unambiguous focal point increases the cooperation rate in a one-shot public goods game. Furthermore, the increase in contributions is associated with a one-to-one increase in beliefs about contributions of other group members. Furthermore, since the increase in contributions is predominantly driven by individuals who would have contributed a positive amount anyway, we suggest that the focal point provides *conditional cooperators* with a means to coordinate their behavior.

Moreover, we present evidence that the opportunity to announce ex post one's contribution decision does not have a significant impact on the cooperation rate. Harnessing guilt/shame and lying aversion therefore proves not to be sufficient for supporting cooperation under low observability. We leave it to future research to investigate how the institutional environment can be changed to increase guilt/shame and lying costs for the benefit of cooperation.

References

Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics 113*, 96–104.

Abeler, J., D. Nosenzo, and C. Raymond (2018). Preferences for truth-telling. *Econometrica*.

Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review 102*(7), 3317–3332.

Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics 110*(1), 1–21.

Andreoni, J. and J. Miller (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica 70*(2), 737–753.

Andreoni, J. and R. Petrie (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics 88*(7), 1605–1623.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*.

Battigalli, P., G. Charness, and M. Dufwenberg (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization 93*, 227–232.

Battigalli, P. and M. Dufwenberg (2007). Guilt in games. *American Economic Review 97*(2), 170–176.

Bhattacharya, P., K. Nielsen, and A. Sengupta (2017). Timing of communication. *Working Paper*.

Bochet, O., T. Page, and L. Putterman (2006). Communication and punishment in voluntary contribution experiments. *Journal of Economic Behavior & Organization 60*(1), 11–26.

Bochet, O. and L. Putterman (2009). Not just babble: Opening the black box of communication in a voluntary contribution experiment. *European Economic Review 53*(3), 309–326.

Bolton, G. E. and A. Ockenfels (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 166–193.

Cantoni, D., D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Protests as strategic games: Experimental evidence from Hong Kong's democracy movement.

Charness, G. and M. Dufwenberg (2006). Promises and partnership. *Econometrica 74*(6), 1579–1601.

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. *Experimental Economics 14*(1), 47–83.

Chen, D. L., M. Schonger, and C. Wickens (2016). oTree - An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review 100*(4), 1358–98.

Childs, J. (2012). Gender differences in lying. *Economics Letters 114*(2), 147–149.

Childs, J. (2013). Personal characteristics and lying: An experimental investigation. *Economics Letters 121*(3), 425–427.

Cinyabuguma, M., T. Page, and L. Putterman (2005). Cooperation under the threat of expulsion in a public goods experiment. *Journal of Public Economics 89*(8), 1421–1435.

Cohn, A., M. A. Maréchal, and T. Noll (2015). Bad boys: How criminal identity salience affects rule violation. *Review of Economic Studies 82*(4), 1289–1308.

Conrads, J., B. Irlenbusch, R. M. Rilke, A. Schielke, and G. Walkowitz (2014). Honesty in tournaments. *Economics Letters 123*(1), 90–93.

Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics 3*(1), 55–79.

Croson, R. and M. Marks (2001). The effect of recommended contributions in the voluntary provision of public goods. *Economic Inquiry 39*(2), 238–249.

Dal Bó, E. and P. Dal Bó (2014). Do the right thing: The effects of moral suasion on cooperation. *Journal of Public Economics 117*, 28–38.

De Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and bounding experimenter demand. *American Economic Review 108*(11), 3266–3302.

Dreber, A. and M. Johannesson (2008). Gender differences in deception. *Economics Letters 99*(1), 197–199.

Dufwenberg, M., S. Gächter, and H. Hennig-Schmidt (2011). The framing of games and the psychology of play. *Games and Economic Behavior 73*(2), 459–478.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Dutta, B. and A. Sen (2012). Nash implementation with partially honest individuals. *Games and Economic Behavior 74*(1), 154–169.

Ellingsen, T. and M. Johannesson (2004). Promises, threats and fairness. *Economic Journal 114*(495), 397–420.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010). Testing guilt aversion. *Games and Economic Behavior 68*(1), 95–107.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives 3*(4), 99–117.

Erat, S. and U. Gneezy (2012). White lies. *Management Science 58*(4), 723–733.

Fehr, E. and S. Gächter (2000). Cooperation and punishment in public goods experiments. *American Economic Review 90*(4), 980–994.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 817–868.

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in disguise - An experimental study on cheating. *Journal of the European Economic Association 11*(3), 525–547.

Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters 71*(3), 397–404.

Frey, B. S. and S. Meier (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review 94*(5), 1717–1722.

Friesen, L. and L. Gangadharan (2012). Individual level evidence of dishonesty and the gender effect. *Economics Letters 117*(3), 624–626.

Gächter, S. and E. Renner (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics 13*(3), 364–377.

Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review 103*(1), 532–548.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 384–394.

Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying aversion and the size of the lie. *American Economic Review 108*(2), 419–53.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association 1*(1), 114–125.

Gunnthorsdottir, A., D. Houser, and K. McCabe (2007). Disposition, history and contributions in public goods experiments. *Journal of Economic Behavior & Organization 62*(2), 304–315.

Gylfason, H. F., A. A. Arnardottir, and K. Kristinsson (2013). More on gender differences in lying. *Economics Letters 119*(1), 94–96.

Hermle, L. H. J. and A. R. C. Roth (2018). Political activists as free-riders: Evidence from a natural field experiment.

Irlenbusch, B. and J. Ter Meer (2013). Fooling the nice guys: Explaining receiver credulity in a public good game with lying and punishment. *Journal of Economic Behavior and Organization 93*, 321–327.

Irlenbusch, B. and J. Ter Meer (2015). Lying in public good games with and without punishment. Technical report, Cologne Graduate School in Management, Economics and Social Sciences.

Isaac, R. M. and J. M. Walker (1988). Communication and free-riding behavior: The voluntary contribution mechanism. *Economic Inquiry 26*(4), 585–608.

Kartik, N., O. Tercieux, and R. Holden (2014). Simple mechanisms and preferences for honesty. *Games and Economic Behavior 83*, 284–290.

Keser, C. and F. Van Winden (2000). Conditional cooperation and voluntary contributions to public goods. *Scandinavian Journal of Economics 102*(1), 23–39.

Kessler, J. B. (2017). Announcements of support and public good provision. *American Economic Review 107*(12), 3760–3787.

Koessler, A.-K., L. Page, and U. Dulleck (2018). Public statements of good conduct promote pro-social behavior.

Krupka, E. L. 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.

Ledyard, J. (1997). Public goods: A survey of experimental research. In J. H. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics*. Priceton University Press.

López-Pérez, R. (2008). Aversion to norm-breaking: A model. *Games and Economic Behavior 64*(1), 237–267.

Masclet, D., C. Noussair, S. Tucker, and M.-C. Villeval (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review*, 366–380.

Matsushima, H. (2008). Role of honesty in full implementation. *Journal of Economic Theory 139*(1), 353–359.

Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review 84*(3), 658–673.

Miettinen, T. (2013). Promises and conventions - An approach to pre-play agreements. *Games and Economic Behavior 80*, 68–84.

Miettinen, T. and S. Suetens (2008). Communication and guilt in a prisoner's dilemma. *Journal of Conflict Resolution 52*(6), 945–960.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper and Row.

Muehlheusser, G., A. Roider, and N. Wallmeier (2015). Gender differences in honesty: Groups versus individuals. *Economics Letters 128*, 25–29.

Neugebauer, T., J. Perote, U. Schmidt, and M. Loos (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology 30*(1), 52–60.

Ortner, J. (2015). Direct implementation with minimally honest individuals. *Games and Economic Behavior 90*, 1–16.

Page, T., L. Putterman, and B. Unel (2005). Voluntary association in public goods experiments: Reciprocity, mimicry and efficiency*. *Economic Journal 115*(506), 1032–1053.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*(5), 1281–1302.

Rege, M. and K. Telle (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics 88*(7), 1625–1644.

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press: Cambridge, MA.

Serra-Garcia, M., E. Van Damme, and J. Potters (2013). Lying about what you know or about what you do? *Journal of the European Economic Association 11*(5), 1204–1229.

Smith, R. H., J. M. Webster, W. G. Parrott, and H. L. Eyre (2002). The role of public exposure in moral and nonmoral shame and guilt. *Journal of Personality and Social Psychology 83*(1), 138.

Tangney, J. P., R. S. Miller, L. Flicker, and D. H. Barlow (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology 70*(6), 1256.

Thaler, R. H. T. and C. R. Sunstein (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Tsikas, S. A., A. Wagener, et al. (2018). Bringing tax avoiders to light: Moral framing and shaming in a public goods experiment. Technical report, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät.

Vanberg, C. (2008). Why do people keep their promises? An experimental test of two explanations. *Econometrica 76*(6), 1467–1480.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

Appendix A: Theoretical Framework

The experimental design described above outlines the way in which we manipulated the decision environment faced by participants along two dimensions. The following section describes our hypotheses regarding the mechanisms through which these treatment manipulations might influence behavior. Importantly, across all our treatments, an agent who cares only about maximizing her extrinsic monetary earnings will always choose a contribution level of zero. Therefore, our treatment manipulations aim to shift only the agent's intrinsic motives for contributing.

In order to express ideas slightly more formally, we will consider a simple model of a participant's decision making in the one-shot public goods game. As above, denote individual i's contribution to the group project by g_i, and let her announcement in the announcement stage be r_i. Individual i's choice vector is (g_i, r_i), where $g_i, r_i \in \{0, \ldots, 20\}$. Let g_{-i} denote the vector of contributions by i's group members. Given a contribution profile (g_i, g_{-i}), individual i's monetary payoff π_i^m is:

$$\pi_i^m(g_i, g_{-i}) = 20 - 0.65 \cdot g_i + 0.35 \cdot \sum_{j \neq i} g_j. \tag{3}$$

As mentioned above, since the individual's payoff is strictly decreasing in her own contribution level, g_i, a monetary payoff maximizing agent will choose a contribution level of zero. Therefore, we augment this simple objective function and consider an agent who may also face several types of intrinsic motives that might influence her choice of contribution level.

Firstly, we assume that the agent obtains intrinsic disutility from contributions that differ from some personal reference contribution level $N_i \in \{0, \ldots, 20\}$. Depending on the agent's motivations, we may think of N_i as representing either (i) her perception of a socially appropriate contribution level (an injunctive social norm), or (ii) her perception of what her peers are likely to contribute (a descriptive social norm).[39]

In case (i), N_i would be a fixed contribution level known to the agent. In case (ii), the agent may be influenced by her belief about the average (denoted by $\bar{g}_{-i}$) contribution level of her group members.[40, 41] Moreover, in case (ii), one may think of the disutility representing a reduced form

[39] One way to think about this is that different individuals construct different mental representations of the decision problem, and for some their subjective injunctive norm is important for guiding their decision, while for others their subjective belief regarding the descriptive norm is important for their decision making. We incorporate both possibilities into our simple model.

[40] One can replace the average contribution level with the modal contribution level and the logic is largely unchanged.

[41] For different approaches of modeling social norms in economics, see e.g. Elster (1989); López-Pérez (2008); Krupka and Weber (2013).

of a reciprocal motive, whereby the agent prefers to match her contribution with the average contribution by her group members (Rabin, 1993). For simplicity, we assume that the disutility takes a quadratic form $-\lambda_i(g_i - N_i)^2$, where $\lambda_i \geq 0$ denotes the weight the agent assigns to the intrinsic component of her utility.[42]

Secondly, following a substantial literature documenting an aversion to lying, we assume that the agent incurs a lying cost, $L_i(g_i, r_i)$, when not reporting her contribution truthfully. This literature demonstrates that many people prefer to tell the truth, even when it is financially costly and there is no risk of being caught, observed or punished (Gneezy, 2005; Fischbacher and Föllmi-Heusi, 2013; Serra-Garcia et al., 2013; Abeler et al., 2014). We abstract away from the literature discussing heterogeneity in lying costs as a function of the 'size of the lie' and assume that there is a fixed cost induced by any report different to the truth, i.e. $r_i \neq g_i$ (Gneezy et al., 2018). However, we allow for heterogeneity in this fixed cost *across* individuals (Gibson et al., 2013). More specifically, we assume that individual i's lying cost takes the following form:

$$L_i(g_i, r_i) = \begin{cases} 0, & r_i = g_i \\ K_i, & r_i \neq g_i, \end{cases}$$

where $K_i \geq 0$.

Finally, we model a cost of announcing a low contribution to one's group members, $A_i(r_i) \geq 0$. We assume that the cost is decreasing as the announcement increases, i.e. $A'(r_i) \leq 0$, and that the cost decrease is smaller when the announcement level is high, i.e. $A_i''(r_i) \geq 0$. Since the announcement is anonymous, $A(\cdot)$ is not a standard social image cost associated with an individual's aversion to others receiving a negative signal that they can link to her identity. Furthermore, since we consider a one-shot contribution game, there is no strategic motive to inflate announcements in order to induce others to contribute more in the future. Instead, we view $A_i(r_i)$ as something similar to a self image cost.[43]

Throughout, we assume that the individual is risk neutral with respect to her monetary payoff, and that the extrinsic and the intrinsic components of her utility are additively separable. Given these assumptions, we can consider an individual's objective function as comprising the sum of

[42] That the disutility can be caused by both negative and positive deviations from N_i may be too strong an assumption if one views N_i as representing an injunctive norm. However, using a function $-\lambda_i(\min\{0, g_i - N_i\})^2$ instead does not change our qualitative results. Cohn et al. (2015) use a similar kind of quadratic loss function to model the disutility a moral agent suffers when deviating from a known moral action. With the interpretation $N_i = \bar{g}_{-i}$, the disutility term is also closely related to disutility an agent suffers when her payoff differs from the payoff of others in the Fehr-Schmidt model (Fehr and Schmidt, 1999).

[43] Equivalently, one can view this term as reflecting an aversion to revisiting a decision that one is ashamed of.

her extrinsic and intrinsic payoffs:

$$U_i(g_i, r_i, g_{-i}; N_i) = \pi_i^m(g_i, g_{-i}) - \lambda_i(g_i - N_i)^2 - A_i(r_i) - L_i(g_i, r_i). \tag{4}$$

As the individual does not know the contributions by her group members ex ante, she maximizes the expectation of 4, where the expectation is taken with respect to a (joint) probability distribution over g_{-i} (denoted by $\mathbb{E}_{g_{-i}}$).[44]

Since the individual's extrinsic monetary payoff function, π_i^m, remains constant across all treatments, and is described in equation 3, we will focus on the difference in her intrinsic payoff components across treatments. Furthermore, for ease of exposition, we assume that g_i and r_i are continuous variables on $[0, 20]$.

The BASELINE treatment

In the BASELINE treatment, there is no possibility to make an ex post announcement, and thus the last two terms in the utility function are set equal to zero. Therefore, the individual trades off her extrinsic monetary motive against the quadratic disutility, in order to choose a contribution level. It is straightforward to see that if $\lambda_i = 0$ (i.e. the agent is selfish), and/or $\mathbb{E}_{g_{-i}}(N_i) = 0$, then the optimal contribution is 0.[45] However, if both λ_i and $\mathbb{E}_{g_{-i}}(N_i)$ are positive, the agent may optimally choose a positive contribution. Note that in order for the right boundary to be optimal, it must be the case that $N_i = 20$ and that λ_i is sufficiently large. Given an interior solution, the necessary condition for the optimum is

$$-0.65 - \frac{\partial}{\partial g_i}\mathbb{E}_{g_{-i}}(\lambda_i(g_i - N_i)^2) = 0.$$

Solving for the optimum gives[46]

$$g_i^* = \mathbb{E}_{g_{-i}}(N_i) - \frac{0.65}{2\lambda_i}. \tag{5}$$

Note that our simple model can accommodate many of the typical qualitative findings in the literature on linear public goods games. In particular, it is a robust finding that most people can be

[44] It is important to note that if N_i is a fixed contribution level, exogenous to the model, as in interpretation (i) above, it holds that $\mathbb{E}_{g_{-i}}(N_i) = N_i$. However, the same is not true if $N_i = \bar{g}_{-i}$, as in interpretation (ii) above.

[45] If $\lambda_i = 0$, function 4 reduces to $U_i(g_i, g_{-i}) = \pi_i^m(g_i, g_{-i})$, the expectation of which is maximized at $g_i = 0$. If $\lambda_i > 0$ and $\mathbb{E}_{g_{-i}}(N_i) = 0$, the expectation of function 4 reduces to $\mathbb{E}_{g_{-i}}(\pi_i^m(g_i, g_{-i})) - \lambda_i(g_i^2 + \mathbb{E}_{g_{-i}}(N_i^2))$, which is clearly maximized by setting $g_i = 0$.

[46] We can interchange the order of partial differentiation and integration. See the measure-theoretic form of the Leibniz Integral Rule.

classified in one of three categories: free-riders ($\lambda_i = 0$ in our model), unconditional cooperators (a fixed $N_i = 20$ and a large λ_i in our model), or self-servingly conditional cooperators (an interior optimum with $N_i = \bar{g}_{-i}$ in our model).[47]

The NUDGE treatment

In the NUDGE treatment, contributions of "16 or greater" are labelled as "good". We posit that this treatment to has an impact on how subjects form their expected reference contribution level, $\mathbb{E}_{g_{-i}}(N_i)$. More specifically, for an agent in case (i) who takes N_i to be an injuctive norm indicating what the appropriate behavior is, we expect the nudge to shift N_i, such that $N_i \geq 16$. Thus, individuals who would have considered it to be socially appropriate to contribute relatively low (e.g. a 50/50 split between the private and the group project) in the absence of the nudged instructions, would increase their contribution when nudged to contribute at least 16.[48]

Furthermore, in the NUDGE treatment, 16 becomes a focal contribution level. Hence, we expect that some individuals shift their belief regarding the average contribution level of others to sixteen, i.e. $\mathbb{E}_{g_{-i}}(\bar{g}_{-i}) = 16$.[49] Thus, for an agent in case (ii) who takes N_i to be the descriptive norm, $N_i = \bar{g}_{-i}$, the nudge may shift her expectation regarding the value of N_i, thereby influencing her contribution choice. For such a conditional cooperator, the contribution choice is a coordination problem that the focal point helps to solve by serving as a natural coordination point. Therefore, conditional cooperators who would have had low beliefs, and thus a low contribution, in the absence of the focal point, should increase their contribution choice to be closer to 16 in the NUDGE treatment.

As can be seen from the expression for the interior optimum in equation 5, an upward shift in $\mathbb{E}_{g_{-i}}(N_i)$ increases contributions. In addition, given that we expect $\mathbb{E}_{g_{-i}}(\bar{g}_{-i}) = 16$ to hold for many conditional cooperators, if their reciprocal motive is strong enough (i.e. a high λ_i), we expect

[47] Empirical evidence on the association between contributions and beliefs about other people's contributions in public good type situations is mixed. Many laboratory and field studies find a positive association (Fischbacher et al., 2001; Frey and Meier, 2004; Chen et al., 2010), while others find a negative association (Cantoni et al., 2017; Hermle and Roth, 2018).

[48] One possible reason for lower assessments of the injuctive norm in the BASELINE treatment is that ex ante subjects may engage in motivated reasoning regarding what is appropriate and choose to hold a low belief in order to justify their own lower contribution. In the NUDGE treatment, there is no longer ambiguity or scope to engage in this type of motivated reasoning regarding what is viewed as appropriate.

[49] Similarly, in a version of the model where individuals care about the modal contribution of their peers, we would expect that the NUDGE treatment shifts the expected modal contribution level to sixteen, i.e. $\mathbb{E}_{g_{-i}}(\textit{Mode}(g_{-i})) = 16$.

many conditional cooperators to contribute 16.[50] Finally, we do not expect the NUDGE treatment to have an impact on the selfish individuals (those with $\lambda_i = 0$). This is summarized in our first hypothesis.

> **Hypothesis 1.** *(i) The average contribution is higher in the* NUDGE *treatment relative to the* BASELINE *treatment. (ii) The fraction of contributions above sixteen is higher and the fraction of intermediate level contributions (i.e. between 1 and 15) is lower in the* NUDGE *treatment relative to the* BASELINE *treatment.*

The ANNOUNCEMENT treatment

In the ANNOUNCEMENT treatment, the agent is potentially influenced by the additional intrinsic motives reflected by the announcement cost for low reports, $A_i(r_i)$, as well as a cost accrued if she falsely reports her contribution level, $L_i(g_i, r_i)$. The relevant objective function is then given by equation 4.

The comparison of the ANNOUNCEMENT and BASELINE treatments allows us to test the hypothesis that knowing that one will later have to announce one's contribution level to the other individuals who are influenced by one's choice might influence one's contribution choice. If true, this would imply that the intrinsic motive not to lie, in conjunction with an intrinsic cost associated with announcing a low contribution level could be used to improve the cooperation rates in public goods games where it is difficult to monitor actual contribution choices, simply by requiring an ex post announcement.[51]

The logic behind this hypothesis is the following. Consider a subject contemplating contributing a low amount to the group project (i.e. someone who would make a low contribution in the BASELINE treatment). She now anticipates that contributing a low amount entails either truthfully revealing her selfishness to her group members in the announcement stage, or lying about her contribution. If the subject is lying averse, the latter announcement strategy means she will suffer the intrinsic lying cost, $L_i(g_i, r_i)$. However, the alternative of truthfully revealing her selfish choice to her group may also entail an intrinsic announcement cost, $A_i(r_i)$. One way to both avoid the

[50] Of course, conditional cooperators (from case (ii)) who would have had a belief *above* 16 in the absence of the focal point, should instead decrease their contribution. Similarly, those in case (i) who would have considered 20 to be the appropriate contribution in the absence of the nudge, should decrease their contribution to 16. However, given the observed contribution distributions in previous literature, we expect the impact on the intermediate contributors to dominate.

[51] While both the aversion to lying and announcement cost motives would likely be stronger if anonymity were relaxed, in order to isolate internal intrinsic self-conscious motives (as opposed to social image concerns), our experimental design maintains anonymous announcements.

lying cost, and alleviate the announcement cost associated with stating a low contribution level is to increase her contribution level.

In order to derive the predictions of our model slightly more formally, we first make four observations concerning our simple model's predictions for behavior in the ANNOUNCE treatment.

Observation 1 An agent who would have contributed 20 in the BASELINE treatment, contributes 20 and announces it truthfully. This is because $g_i = 20$ maximizes the expectation of expression $\pi_i^m - \lambda_i(g_i - N_i)^2$, and given $g_i = 20$, truthful announcement minimizes both $L_i(g_i, r_i)$ and $A_i(r_i)$.

Observation 2 For a given contribution level, g_i, choice vectors (g_i, r_i) with $g_i \neq r_i$ and $r_i < 20$ are dominated by the vector $(g_i, 20)$.[52] Hence an agent who lies (i.e. has a low K_i), chooses the vector $(g_i^{BASE}, 20)$, where $g_i^{BASE} < 20$ is the contribution that would have been her optimal choice in the BASELINE treatment.

Observation 3 An agent who would have contributed a positive amount g_i^* in the BASELINE either contributes g_i^* and lies maximally (if K_i is low), or is truthful (if K_i is high) and contributes either 20 or $g_i^{**} > 0$, where g_i^{**} solves the first-order condition

$$\frac{\partial \mathbb{E}_{g_{-i}}(U_i(g_i, r_i, g_{-i}; N_i))}{\partial g_i} = -0.65 - 2\lambda_i g_i + 2\lambda_i \mathbb{E}_{g_{-i}}(N_i) - A'(g_i) = 0.$$

Additionally, it holds that $g_i^{**} \geq g_i^*$ since

$$-0.65 - 2\lambda_i g_i^* + 2\lambda_i \mathbb{E}_{g_{-i}}(N_i) - A'(g_i^*) \geq 0,$$

and since both $-2\lambda_i g_i$ and $-A'(g_i)$ are decreasing in g_i.

Observation 4 An agent who would have contributed 0 in the BASELINE treatment either contributes 0 and announces it truthfully (if A_i low and K_i is high), contributes 0 and announces 20 (if A_i high and K_i is low), or contributes a positive amount and announces it truthfully (if both A_i and K_i are high).

It is clear from observations 1-4 that all agents contribute at least as much in the ANNOUNCE treatment as they would have contributed in the BASELINE treatment. Additionally, some agents may contribute strictly more. This is the basis for our next hypothesis:

[52] In other words, if an agent lies, she lies maximally.

Hypothesis 2. *The average contribution in the* ANNOUNCE *treatment is higher than the average contribution in the* BASELINE *treatment.*

The NUDGE + ANNOUNCE treatment

Finally, in our NUDGE + ANNOUNCE treatment, we examine whether there is an interaction effect between the intrinsic motives discussed above. In particular, one might think that raising the salience of high contributions being socially efficient will also magnify the announcement cost $A_i(r_i)$ for values below sixteen. The intuition is that when one chooses a contribution that is labelled as bad, it is more costly to announce it. If this is true, it would lead to a discontinuity in the announcement cost function at 16. Therefore, the objective function in this final treatment can be represented as follows.

$$U_i(g_i, r_i, g_{-i}; N_i) = \pi_i^m(g_i, g_{-i}) - \lambda_i(g_i - N_i)^2 - A_i(r_i) \cdot [1 + \theta_i \cdot \mathbb{1}(r_i < 16)] - L_i(g_i, r_i), \quad (6)$$

where $\theta_i \geq 0$.

Notice that the (expected) payoff difference between the NUDGE + ANNOUNCE and the NUDGE treatment for strategy (g_i, r_i), given a fixed $\mathbb{E}_{g_{-i}}(N_i)$, is

$$-A_i(r_i) \cdot [1 + \theta \cdot \mathbb{1}(r_i < 16)] - L_i(g_i, r_i), \quad (7)$$

while the (expected) payoff difference between ANNOUNCE and BASELINE for strategy (g_i, r_i), given a fixed $\mathbb{E}_{g_{-i}}(N_i)$, is

$$-A_i(r_i) - L_i(g_i, r_i). \quad (8)$$

Clearly, for any *fixed* strategy (g_i, r_i), expression 7 is less than or equal to expression 8, and for any strategy (g_i, r_i) such that $r_i < 16$, expression 7 is strictly less than expression 8. This would imply a stronger motive to contribute above sixteen when the announcement mechanism is introduced to the NUDGE treatment than when it is introduced to the BASELINE treatment.

On the other hand, it may not be appropriate to compare expression 7 to expression 8 for a *fixed* strategy (g_i, r_i), because contributions could in general be very different in treatments NUDGE and BASELINE. If the NUDGE treatment is successful in shifting the majority of contributions to be above 16 through inducing a higher $\mathbb{E}_{g_{-i}}(N_i)$, there might not be much room for improvement in the NUDGE + ANNOUNCE. Therefore, introducing the announcement mechanism may in fact have a greater positive impact on contributions in the BASELINE treatment compared to the NUDGE treatment. We summarize these mutually exclusive possibilities in the following two hypotheses:

Hypothesis 3a. *The increase in average contribution in the* NUDGE + ANNOUNCE *treatment relative to the* NUDGE *treatment is larger than the increase in average contribution in the* ANNOUNCE *treatment relative to the* BASELINE *treatment.*

Hypothesis 3b. *The increase in average contribution in the* ANNOUNCE *treatment relative to the* BASELINE *treatment is larger than the increase in average contribution in the* NUDGE + ANNOUNCE *treatment relative to the* NUDGE *treatment.*

In the scenario where the NUDGE treatment alone has a small effect relative to the BASELINE, hypothesis 3a is the appropriate one. However, in the scenario where the NUDGE treatment alone has a large effect relative to the BASELINE, hypothesis 3b is the appropriate one.

Appendix B

Figure 4: CDF of contribution levels across treatments

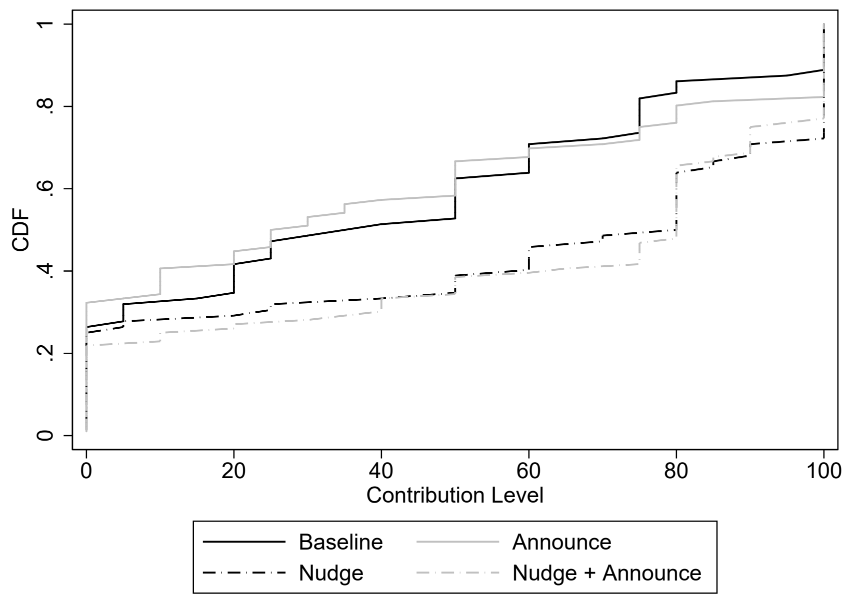


Figure 5: Distribution of contribution levels

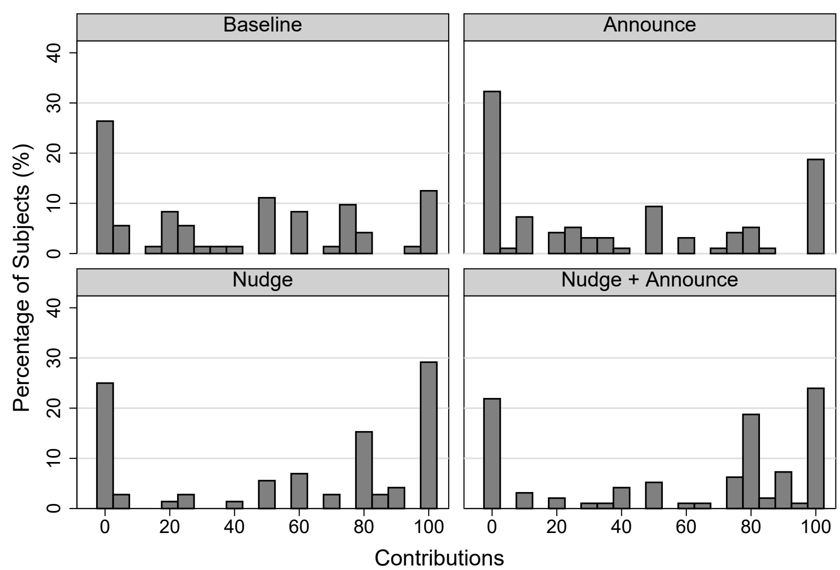


Figure 6: Ex ante beliefs across treatments (mean and median)

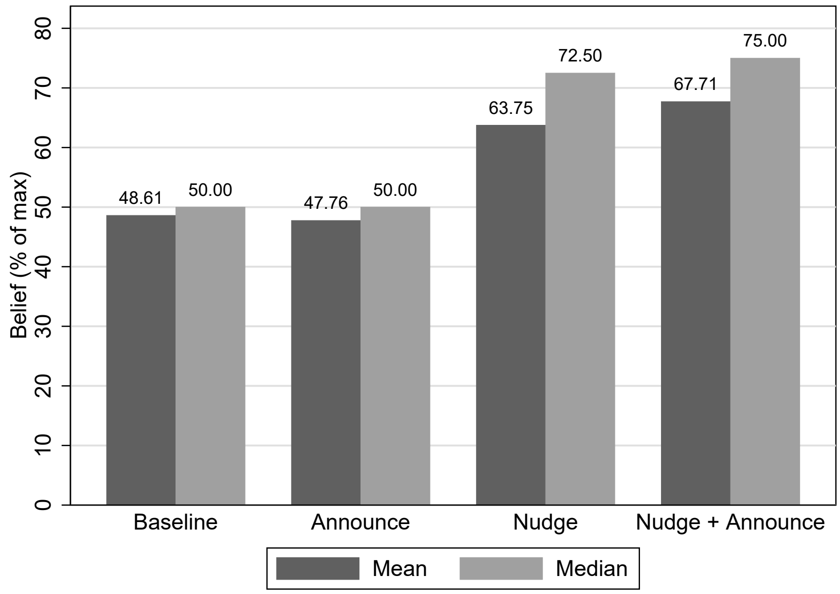


Figure 7: Beliefs and contributions

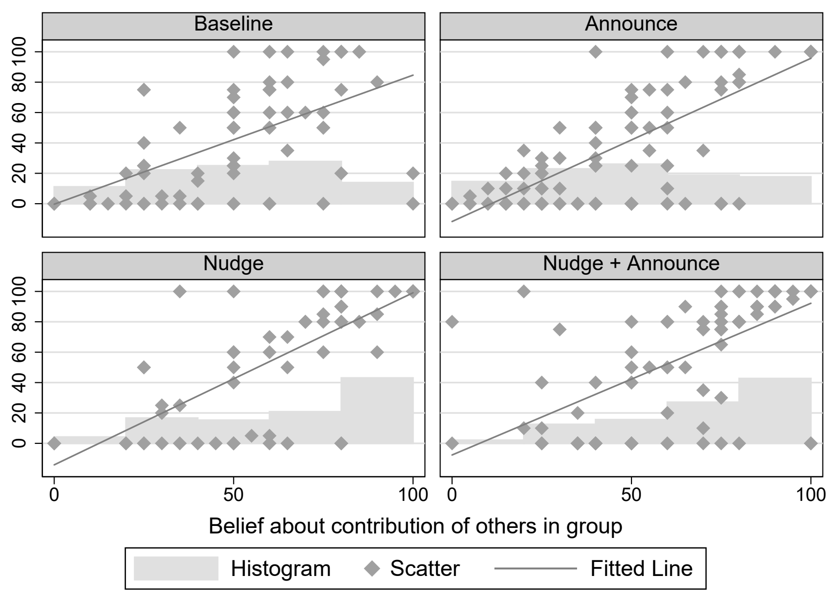


Table 4: Tobit Estimates

	Contributions (1a)	Ex Ante Beliefs (1b)	Contributions (2a)	Contributions (2b)
Treatments				
Nudge ($T_N = 1$)	27.0**	16.6***		-0.7
	(11.96)	(4.85)		(8.56)
Announcement ($T_A = 1$)	-1.9	-0.04		1.8
	(11.14)	(4.52)		(7.86)
Nudge * Announcement ($T_N * T_A = 1$)	2.9	4.4		-6.4
	(15.81)	(5.72)		(11.17)
Ex ante belief			1.7***	1.8***
			(0.12)	(0.13)
Constant	34.0***	48.3***	-50.6***	-51.4**
	(8.38)	(3.41)	(7.46)	(8.84)
Observations	336	336	336	336
Left-censored	89	12	89	89
Right-censored	71	28	71	71

Notes: (i) Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$)

(ii) The regression includes a binary dummy variable for the nudge treatment, a binary dummy variable for the announce treatment, and a dummy for the interaction effect (as opposed to a dummy variable for each treatment).

Appendix C

Instructions for the NUDGE + ANNOUNCE *treatment.*

Game II: Overview

In Game II, all participants are divided into groups of four. These are completely new groups and are not related to the groups from Game I in any way. (Recall that Game II is not related in any way to Game I). You will therefore be in a group with three other participants (called partners later on). However, you will not learn the identities of your partners, nor will they learn yours.

Game II of the experiment consists of two stages. At the beginning of the first stage you (and each of your partners) will receive an endowment of 5 EUR. This endowment will be converted into experimental points using the following exchange rate:

4 points = 1 EUR

Therefore, you will start Game II of the experiment an endowment of 20 points. At the end of the experiment, we will convert your points back to Euro using the same exchange rate as indicated above and it will be added to your earnings from the rest of the experiment and paid to you in cash.

The Contribution Stage:

In the contribution stage, you will need to decide how many of the 20 points you want to contribute to a group project, and how many you would like to keep yourself. Every point that you keep yourself, increases your final earnings by 1 point. Every point that you contribute to the group project will increase your own earnings by 0.35 points, and will also increase the earnings of each member of your group by 0.35 points. Similarly, every point that another group member contributes to the group project will also increase your earnings by 0.35 points. At the end of the experiment, you will learn the value of your earnings, but the amount contributed by each participant will not be revealed.

Your payoff from the group project will be calculated as follows.

First, all the group contributions in your group are summed up. You (and each of your partners) then get 0.35 * the sum of your group contributions. This will be added to the amount that you kept for yourself.

Hence your total payoff from the Contribution Stage consists of two components:

1) the fraction of your endowment you have kept for yourself,

2) your share of the total group project contributions.

Your earnings can therefore be calculated as follows:

Earnings = (20 - your group contribution) + (0.35 * total group contribution)

Examples:

1. Imagine that you and all your group members contribute the full endowment (i.e. 20 points); the sum of group contributions in your group is 80 points. You would get 0.35 * 80 = 28 points from the group project. Hence your total payoff is (20 - 20) + 28 = 28 points. This is also what your partners get.

2. Imagine that you and all your group members contribute nothing to the group project (i.e. 0 points); the sum of group contributions in your group is 0. You get 0.35 * 0 = 0 points from the group project. Hence your total payoff is (20 - 0) + 0 = 20 points. This is also what your partners get.

Before making your contribution decisions, we will ask you some questions on the computer to ensure that you have fully understood these instructions. In addition, during this initial clarification phase, there will be a calculator on the screen that allows you to input hypothetical contribution values for you and all your group members and the calculator will tell you how much each person would earn.

Contributing to the group project is beneficial for everyone

As you can see from the examples above, if everyone contributes to the group project, you will all earn more points than if you all contribute nothing and keep all your points to yourself. It is therefore beneficial to everyone to contribute more to the group project.

Of course, if everyone else contributes to the group project and you keep all your points to yourself, then you will earn even more, but this will harm the other members of your group. They will lose more than you gain.

A contribution of 16 or above is "good"

While you may choose any contribution level you would like, in this experiment, we will call a contribution of 16 or larger a "good" contribution, and a contribution of 15 or below a "bad" contribution, since contributing to the group is good for everyone. On your decision screen, you will see that the "good" contribution values are denoted in green, while the "bad" contribution values are denoted in red.

The Announcement Stage:

After you have made your contribution decision, every member of the group will make an announcement to the other members of the group regarding the amount that he / she has contributed. Your final payment will depend on the contribution you made and not on your announcement.

The announcement will take the following form and you will have to choose what message to send in place of x and 20-x :

"I contributed x points to the group project and kept 20-x points for myself"

Once all the group members have made their announcements, you will observe the announced contributions of your group members.

We will now proceed to Game II. Before we do, if you have any questions at this moment, please raise your hand. The experimenter will come to you.

Discussion Papers of the Research Area Markets and Choice 2018

Research Unit: **Market Behavior**

Mustafa Oğuz Afacan, Inácio Bó, Bertan Turhan Assignment maximization	SP II 2018-201
Inácio Bó, Chiu Yu Ko Competitive screening and information transmission	SP II 2018-202
Dirk Engelmann, Jana Friedrichsen, Dorothea Kübler Fairness in markets and market experiments	SP II 2018-203
Daphne Chang, Roy Chen, Erin Krupka Rhetoric matters: A social norms explanation for the anomaly of framing	SP II 2018-204
Christian Basteck, Bettina Klaus, Dorothea Kübler How lotteries in school choice help to level the playing field	SP II 2018-205
Mira Fischer, Valentin Wagner Effects of timing and reference frame of feedback: Evidence from a field experiment	SPII2018-206
Hande Erkut Social norms and preferences for generosity are domain dependent	SPII2018-207
Homayoon Moradi Selfless ignorance: Too good to be true	SPII2018-208

Research Unit: **Economics of Change**

Kai Barron, Christina Gravert Confidence and career choices: An experiment	SP II 2018-301
Piotr Evdokimov, Umberto Garfagnini Communication and behavior in organizations: An experiment	SP II 2018-302
Andrea Contigiani, David H. Hsu, Iwan Barankay Trade secrets and innovation: Evidence from the “inevitable disclosure” doctrine	SP II 2018-303
Thomas Mariotti, Nikolaus Schweizer, Nora Szech, Jonas von Wangenheim Information nudges and self-control	SP II 2018-304
Kai Barron, Tuomas Nurminen Nudging cooperation	SP II 2018-305

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice

Research Professorship: **Market Design: Theory and Pragmatics**

Daniel Friedman, Sameh Habib, Duncan James, Sean Crockett SP II 2018-501
Varieties of risk elicitation

Junior Research Group: **Neuroeconomics**

Martin Fochmann, Frank Hechtner, Peter Mohr SP II 2018-601
When happy people make society unhappy: How incidental emotions affect compliance behavior

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice