University of Neuchatel

Institute of Economic Research

IRENE Working paper 18-06

When within- and outgroup norms conflict: A public good experiment with strategic ignorance of social norms

Serhiy Kandul

Bruno Lanz

UNIVERSITÉ DE NEUCHÂTEL

Institut de recherches économiques

Noname manuscript No.
(will be inserted by the editor)

When within- and outgroup norms conflict: A public good experiment with strategic ignorance of social norms

Serhiy Kandul · Bruno Lanz

This version: June 2019

Abstract Social norm feedback, i.e. informing people about the behavior of others, has been shown to influence prosocial behavior in many domains, including tax compliance and energy conservation. We introduce social norm feedback in a public good setting and study the interplay between payoff-relevant within-group norms and payoff-irrelevant outgroup norms. We show that conflict between within- and outgroup norms dampens within-group conditional cooperation. Further, participants strategically ignore outgroup norms when these go against self-interest, instead consulting norm information that allows them reducing their contributions. On aggregate, such information acquisition/avoidance strategy favors exposition to norms that hastens the breakdown of cooperation. Finally, norm avoidance is higher when feedback is based on individual rather than group-level comparisons, which is consistent with a self-image cost associated with social norm feedback.

Keywords: Social norms; Deliberate ignorance; Public good game; Self-image concerns; Prosocial behavior; Externalities.

JEL Codes: C91, D12, D62, D91, H41, Q41.

We thank Cristina Cattaneo, Sylvain Chabé-Ferret, Yu Gao, Martin Kocher, Yohei Mitani, Michael Puntiroli, Tobias Regner, Evert Reins, Hiro Sakamoto, Matthias Uhl as well as seminar participants in Kyoto University and IRME workshop for useful comments and discussions. Excellent research assistance was provided by Fanol Kurteshi. Financial support from the Swiss National Science Foundation under grant PYAPP1_173650 is gratefully acknowledged. This research is part of the activities of SCCER CREST (Swiss Competence Center for Energy Research), which is financially supported by the Swiss Commission for Technology and Innovation (CTI) under Grant No. KTI. 1155000154. Any errors are ours.

Serhiy Kandul
University of Neuchâtel, Department of Economics and Business, Switzerland. Mail: A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland; Tel.: +41 32 718 15 96
E-mail: serhiy.kandul@unine.ch

Bruno Lanz
University of Neuchâtel, Department of Economics and Business, Switzerland. Mail: A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland; Tel: +41 32 718 14 55
bruno.lanz@unine.ch

1 Introduction

Social norms are increasingly used as a tool to enhance socially desirable behavior. This class of informational interventions consists in informing people about their behavior relative to the typical (e.g. average) behavior of their reference group. As theories of social norms and general feedback interventions suggest (Elster, 1989; Kluger and DeNisi, 1996; Karlin et al., 2015), people with a preference to conform with norms who learn that their behavior is different from the reference would adjust their behavior without the need for monetary incentives. Such feedback strategies have been exploited in a variety of domains, from reducing electricity and water consumption to increasing tax compliance and changing nutrition habits (Allcott, 2011; Ferraro et al., 2011; Ayers et al., 2013; Costa and Kahn, 2013; Allcott and Rogers, 2014; Robinson et al., 2014; Hallsworth et al., 2017).

We extend the work on social norms in two important dimensions. First, we introduce a repeated public good setting and distinguish between pay-off relevant *within-group* contribution norms and pay-off irrelevant *outgroup* contribution norms. This distinction allows us to investigate the effectiveness of purely informational social norms in the presence of payoff interdependent within-group interaction.

In our baseline condition, participants learn their payoff after every round can therefore infer how their contribution compares to the contributions of the other group members, i.e. their compliance with the within-group contribution norm. If participants care about the contribution of their group members as conditional cooperation postulates (Fischbacher et al., 2001; Fischbacher and Gachter, 2010), they should adjust their behavior towards the within-group contribution norm. In treatments participants also learn their payoff at the end of each round, which allows them to assess the within-group contribution norm, and in addition they receive a payoff-irrelevant feedback about how their contribution compares to the average contribution in other groups in a session. This allows them to infer an outgroup contribution norm.[1] Therefore, participants in the treatments encounter both within- and outgroup contribution norms.

The second main contribution of our work is to differentiate between individual-level vs. group-level feedback, and document the importance of self-image and information avoidance in how participants consider individual-level feedback. In particular, we argue that negative individual-level feedback typically employed in the field (i.e. learning that one is performing individually worse than others) imposes self-image costs and may therefore prompt people to ignore this information (Dana et al., 2007; Mazar et al., 2008; Matthey and Regner, 2011; Grossman and van der Weele, 2017). The importance of infor-

[1] We inform participants whether their contribution is below, above or, exactly at the average of contributions by all other participants in the session that are not in their group. See Section 3 for details.

mation avoidance as an adaptation strategy against social comparison has been recently recognized in theory on social norms (Bicchieri and Dimant, 2019).[2]

In our public good experiment, we introduce a treatment with group-level feedback on outgroup contribution norms which compares total contribution of a group with average total contributions by the other groups in the same session.[3] To quantify outgroup norms avoidance, we investigate treatments with exogenous and endogenous feedback. In treatments with exogenous feedback, all participants receive individual or group-level feedback on outgroup norms every round just before they get the information about their payoff. In treatments with endogenous feedback, after the contribution choice has been made, participants choose whether to reveal their social comparison (i.e to know whether their individual or total group contribution is below, above or equal to average contribution outside their group). Since the group-level feedback does not directly allow for individual comparison of contributions, such feedback is less likely to be avoided by people with self-image concerns.[4]

While the use of group-level feedback may increase self-exposure to the norm, high group cohesiveness is a prerequisite to affect behavior in the same way individual feedback does (Tajfel and Turner, 1986). As studies in social psychology show, in strongly cohesive groups, people who care about the social image of their group are ready to compensate the wrongdoing of their group members (Lickel et al., 2005; Welten et al., 2012). We rely on student population as naturally occurring group and expect some degree of cohesiveness, so that participants in the experiment are expected to care about the standing of their groups in the experiment. Moreover, the example of other more successful groups could simply make high cooperation salient, i.e. the feedback on the contribution of other groups can enhance within-group contributions beyond group social-image channel (Engel et al., 2011; Boehm and Rockenbach, 2013).[5]

Importantly, social norm feedback in the field is typically coupled with a reference to injunctive norms to avoid a boomerang effect, the downward adjustment to the norm by good performing individuals (e.g. via a signal of social approval with emoticon, see Schultz et al., 2007, for example).[6] Contributions in field studies therefore also leverage moral suasion (see also Ito et al., 2018). Our feedback mechanism

[2] Anecdotal field evidence from Allcott and Kessler (2015) also suggests that 34% of participants in a free social comparison program on energy use hold (weakly) negative willingness to pay for participation in one additional year of the program. According to our work, at least part of observed unwillingness to participate in typical intervention programs comes from self-image concerns.

[3] Here we inform participants whether the total contribution of their group is below, above or, exactly at the average total contribution of other groups. See Section 3 for details. In the field, this could for instance take the form of a comparison between energy/water use in a building or city block and average usage in other buildings or blocks in the area.

[4] By the same argument, group-level feedback is expected to be less affected by other adverse properties of individual level feedback intervention such as "rebound effect" (Jacobsen et al., 2012; Harding and Rapson, 2017) or moral licensing (Tiefenbeck et al., 2013; Ho et al., 2016). These potential additional adverse effects of the feedback is not a focus of this paper.

[5] Note again that by comparing group-level and individual-level feedback we focus on informational incentives and depart from discussions about monetary incentives schemes or strategic interactions between individuals and groups (see e.g. Bandiera et al., 2012; Sutter et al., 2013; Babcock et al., 2015; Ladley et al., 2015; Eckel et al., 2016). This is because interventions based on social norms are typically not linked to monetary incentives.

[6] Interesting for the discussion about self-image concerns of individual feedback information, Allcott (2011, p. 1083, emphasis added) notes that "The "Great" group receives two "smiley face" emoticons, the "Good" group receives one, and the "Below Average" group initially received "frown faces" *until customer complaints* ended this practice."

excludes any cues as to the interpretation of information and focuses exclusively on descriptive social norms (Bicchieri and Xiao, 2009; Bicchieri and Dimant, 2019). We do not include any injunctive norms or moral suasion in the design for two reasons. First, social approval might be perceived differently at the individual vs. group level, producing a confound. Second, studies on within-group contributions norms and conditional cooperation typically rely on descriptive norms. Therefore, introduction of the injuctive norms would introduce an unfortunate departure from this literature.

Our results can be summarized as follows. First, we replicate the well-known pattern of conditional cooperation whereby participating contributing below/above within-group average (relatively high/low payoff) tend to increase/decrease their contributions in subsequent rounds, and vice-versa. Second, we provide evidence that the feedback about outgroup contribution norms affects within-group conditional cooperation in two important ways: negative (below average) outgroup social norm feedback mitigates the downwards adjustment of contributions by high within-group contributors, whereas positive (above average) outgroup social norm feedback breaks down the upward adjustment of low within-group contributors. This result is stronger for individual-level feedback, but it is also present for group-level feedback. For both treatments these two countervailing effects cancel each other out on aggregate, leading to similar aggregate contributions across the conditions. These findings suggest that the feedback on outgroup social norms alone is not sufficient to increase contributions in a public good game setting. Our results also confirm the presence of a boomerang effect in individual feedback treatments, whereas this effect is less pronounced in treatments using group-level comparison.

Second, by introducing an endogenous social feedback mechanism in a public good game, we show that information avoidance in relation to individual feedback is important. At the end of the last round, 37% of participant in the condition with endogenous individual feedback on outgroup contribution norms chose not to consult the feedback. By contrast, when we introduce a social comparison mechanism that relies on group-level behavior, only 10% of participant in this condition choose not to consult the feedback. This suggests that group-level feedback is less prone to information avoidance, with the likelihood of ignoring social comparison feedback being on average about three times smaller than in the individual-level comparison condition. Lower self-exposure to individual feedback interventions can be interpreted as self-image cost, and field implementations of social feedback may therefore benefit from designs that minimize these costs.

The remainder of the paper is organized as follows. Section 2 lays out a simple theoretical framework linking public good contributions, preferences for compliance with norms, and self-image concerns. In Section 3, we describe our public good game and alternative experimental conditions. Section 4 summarizes our results. Section 5 provides some discussion and concluding comments.

2 Public good contributions, norm compliance, and self-image

This section considers a simple theoretical framework that rationalizes how feedback on contributions by others can affect own behavior. Doing so, our main objective is to motivate the experimental design documented below. We start by developing some notation, extending the work of Krupka and Weber (2009) and Gaechter et al. (2017) who both consider how social norms dictated by the behavior of others enter a utility-theoretic framework. We then explain our main expectations regarding the interplay between within-group and outgroup contribution norms. Finally we discuss expected behavior with respect to feedback acquisition/avoidance.

2.1 Utility from norm compliance

We consider a decision-maker i that is part of a group g, as well as a set of groups G, whereby each individual is assigned to a single group. The set of individuals in group g is denoted by i_g. Individual action a_i (contribution to a public good) determines own monetary payoff m_i, together with the behavior (contributions) by other members $-i_g$ that are part of own group g. As in Krupka and Weber (2009) and Gaechter et al. (2017), we assume that decision-maker i forms expectations about the behavior of within-group members, and denote these as $\overline{a}_{-i_g}$. Average contribution in group g is denoted by $a_g(a_i)$. Following the same notational logic, we denote expectations about average contributions by outgroup participants (excluding i and all other members of group g) by $\overline{a}_{-g}$.

We assume decision-maker i selects a_i by maximizing a utility function that depends on m_i and on preferences for compliance with norms:

$$U_i = m_i(a_i) + \beta \cdot N_w[a_i|\overline{a}_{-i_g}] + I \cdot \sigma \cdot N_o[a_i|\overline{a}_{-g}] + I \cdot \gamma \cdot N_g[a_g(a_i)|\overline{a}_{-g}]\,. \tag{1}$$

Preferences for norm compliance are defined at three possible levels: within-group individual comparison (associated with cooperation and reciprocity), individual comparison with outgroup participants, and group-level comparison with other groups. First, $N_w[a_i|\overline{a}_{-i_g}]$ measures the degree of compliance of own action a_i with the behavior of other *within-group* members. In a public good game setting, information about within-group relative standing is typically available through payoff information after each round, and is the main driver behind conditional cooperation motives (Fischbacher et al., 2001; Fischbacher and Gachter, 2010).

Second, $N_o[a_i|\overline{a}_{-g}]$ compares *own individual* contributions a_i and average contributions by members of other groups, i.e. measures the degree of individual compliance with the outgroup contribution norm. Importantly, $\overline{a}_{-g}$ is payoff irrelevant for individual i, so that variation in $\overline{a}_{-g}$ does not trigger cooperation

or reciprocity motives among members of group g. Because it involves individual contribution a_i, we refer to the information contained in $N_o[a_i|\overline{a}_{-g}]$ as "individual feedback" information.

Third, $N_g[a_g(a_i)|\overline{a}_{-g}]$ compares *own-group* total contributions to average total contributions in other groups, i.e. measures the degree of a group compliance with the outgroup contribution norm. Here, individual behavior only affects the distance to the norm through $a_g(a_i)$, and we refer to information about relative standing of own group to other groups as "group-level feedback" information.

The parameters β, σ and γ are the sensitivity parameters to interpersonal comparison $N_w[\cdot]$, $N_o[\cdot]$ and $N_g[\cdot]$ respectively. A parameter I is a dummy indicating whether the information on the outgroup contribution norm is provided. In the baseline, $I = 0$, in the treatments with exogenous feedback $I = 1$; in the treatment with the endogenous feedback I takes the value of 0 or 1 depending on participant's choice.

A preference for norm compliance requires preference parameters β, σ and γ to be positive. Following Krupka and Weber (2009) and Gaechter et al. (2017), we assume that the functions measuring norm compliance admit a unique maximum at the point where individual or group behavior is consistent with the relevant norm. It implies that $N()$ is pushing those below the norm to increase contributions, and those above the norm to decrease their contributions. Therefore, if information about these norms is available we expect that it will affect contributions.

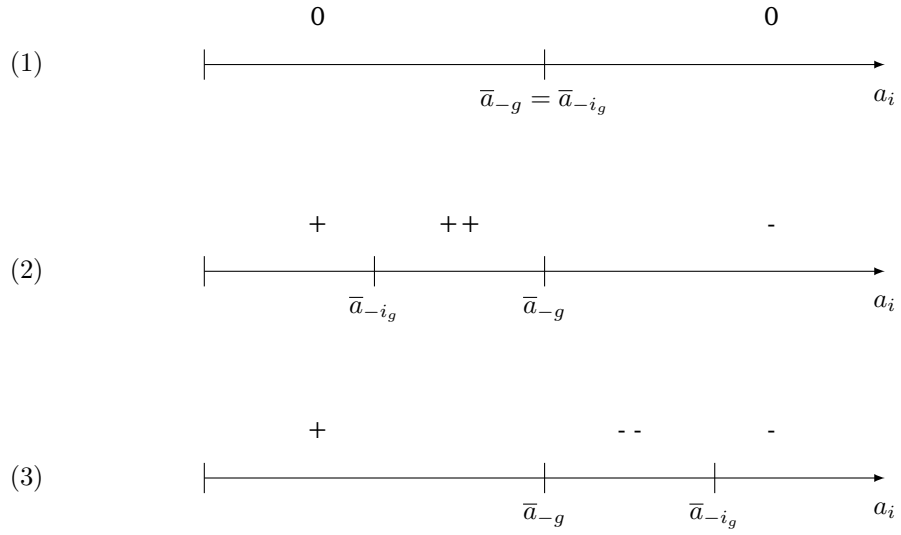
In our experiment, we make the conventional assumption that $\frac{\partial m_i}{\partial a_i} < 0$, which implies that contributions to public good reduce individual welfare all else equal, and puts i in a social dilemma situation with the individually optimal strategy of zero contributions. Since we provide feedback on own payoff in each round, within-group norm $\overline{a}_{-i_g}$ is present in all the treatments including the baseline. What we vary instead is the availability of either individual or group-level feedback on outgroup contribution norms. Therefore, our main empirical strategy is to identify the additional effect of the outgroup social norms, N_o and N_g to the main effect of within-group contribution norms captured by N_w.

2.2 Outgroup contribution norms and conditional cooperation

To understand the within-group cooperation in the absence of any external outgroup comparisons, we define two types of individuals: payoff maximizers and conditional cooperators. In our conceptual framework, payoff maximizers are those with $\beta = 0$, i.e those with no concern about the contribution of their group members. Given $\frac{\partial m_i}{\partial a_i} < 0$, the optimal strategy for these people is always free-ride and contribute nothing to the public good.

As documented by Fischbacher et al. (2001); Fischbacher and Gachter (2010), however, the majority of subjects are typically conditional cooperators, i.e. those who (imperfectly) adjust their contributions to the contributions of their group members. In our framework, this behavior is captured by the positive

Fig. 1: Additional effect of outgroup norms on contributions



Notes: We use +/- and ++/- - to denote respectively small and large increase/decrease in contributions relative to baseline.

parameter $\beta > 0$. High enough β makes those who contribute below the within-group average increase contributions, and those who contribute more than the within-group average decrease contributions. Interestingly, the same towards-the-norm adjustment pattern is observed for payoff-irrelevant comparisons in the social norms literature (Schultz et al., 2007).

To understand how the outgroup contribution norms interacts with within-group cooperation, consider three possible configurations, also depicted in Figure 1:

(1) $\overline{a}_{-i_g} = \overline{a}_{-g}$ - within-group average contribution is the same as the outgroup averagee;

(2) $\overline{a}_{-i_g} < \overline{a}_{-g}$ - within-group average contribution is lower than the outgroup average;

(3) $\overline{a}_{-i_g} > \overline{a}_{-g}$ - within-group average contribution is higher than the outgroup average.

In case (1), within-group and outgroup norms coincide, therefore there is no additional effect of the outgroup contribution norm on contributions. In case (2), when individual contribution is below both within-group and outgroup average ($a_i < \overline{a}_{-i_g} < \overline{a}_{-g}$), preferences for conformism with social norms would trigger an increase of contributions. When individual contribution falls in the range $\overline{a}_{-i_g} < a_i < \overline{a}_{-g}$, however, the two norms are in conflict: a within-group norm signals to decrease contributions, whereas an outgroup norm signals to improve. In the range $a_i > \overline{a}_{-g}$, again both norms suggest to decrease contributions. Similarly, in case (3) in the range $\overline{a}_{-g} < a_i < \overline{a}_{-i_g}$, the two norms are in conflict: an information about the outgroup norm moves the contributions downwards, whereas the within-group norm pushes the contributions upwards; the difference to the baseline with the within-group norm only should be more pronounced.

These three possibilities are defined for both individual and group-level comparisons. By construction, individual feedback $N_o[a_i|\overline{a}_{-g}]$ uses own individual contributions for social comparison, whereas group-

level feedback $N_g[a_g(a_i)|\overline{a}_{-g}]$ own individual contributions affect within-group contributions through $a_g(a_i)$ only. If sensitivity towards individual-level feedback is greater than group-level feedback ($\gamma < \sigma$), individual feedback induces a larger change in contributions relative to group-level feedback, both above and below the norm.

We summarize the hypotheses with respect to the effect of social comparison feedback on contributions as follows:

Hypothesis 1. *Outgroup norms have stronger positive/negative effect on contributions for contributors above/below the within-group average.*

In highly cooperative groups with high within-group average contribution we expect a positive effect of social norm feedback: those above the within-group average who would otherwise be tempted to decreased their contributions are pulled back by the need to meet the outgroup contribution norm. By contrast, in non-cooperative groups with low within-group average contribution, we expect a negative effect of outgroup social feedback: those below the within-group average who would be otherwise tempted to improve towards the within-group average are pulled down since they meet the outgroup contribution norm. If the former effect is stronger, outgroup norms increase contributions. If the latter effect dominates, however, providing information on the outgroup average contribution might backfire.

Hypothesis 2. *Outgroup norms have stronger effect on contributions for individual level feedback.*

Group-level feedback on the outgroup contribution norm mitigates the negative effect of individual-level comparison, but this comes at a cost of potentially smaller positive effect of negative social norm feedback due to lower norm sensitivity: $\gamma < \sigma$.

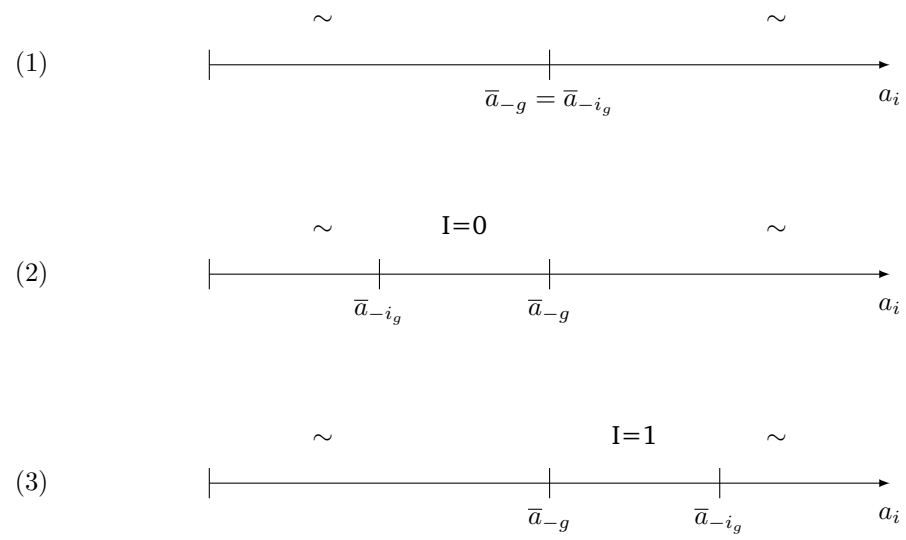
2.3 Outgroup contribution norms and information avoidance

Regarding the information acquisition/avoidance decisions, we follow Grossman and van der Weele (2017) and assume that the self-image costs occur when participants learn they are falling behind the norm. When participants expect their contribution to be below the average contribution in other groups, they suffer a utility loss which is captured by the sensitivity parameters σ and γ. If $\sigma = 0$ and $\gamma = 0$, people are indifferent between ignoring and acquiring the feedback, providing that acquisition is costless. When σ and γ are positive, and the individual is below the norm, preferences to conform with norms would incentivize an increase of contributions.

However, this goes against monetary self-interest, and such a conflict may induce participants to avoid a feedback which damages their self-image and to set $I = 0$.[7] In turn, those who ignore the feedback

[7] Note there might also be a self-image cost associated with non-disclosure of social norm, which could be related to the negative feeling of not daring to obtain self-image relevant information. As studies on deliberate ignorance implicitly show, however, this cost is likely to be negligible (see e.g. Dana et al., 2007).

Fig. 2: Avoidance of information on outgroup contribution norms



Notes: $\sim$ denotes indifference between ignoring the feedback or acquiring it, I=0 denotes the choice to ignore feedback on outgroup contribution norm, and I=1 denotes the choice to acquire the same.

would keep contributing less and earning more. By contrast, when individual contribution is above the norm, moving towards the norm is aligned with monetary self-interest, in this case there is no incentive to ignore the feedback.

Consider again three possible relations between within- and outgroup norms $\overline{a}_{-i_g}$ and $\overline{a}_{-g}$, depicted in Figure 2. In case (1), where the within- and outgroup contribution norms coincide, the information about the outgroup norms does not add any further image costs/benefits. Therefore, people are indifferent between receiving and avoiding the feedback. In case (2), where the within-group norm is below the outgroup one, the contribution range presents a conflict: outgroup norm prescribes to increase contributions, whereas the self-interest and the within-group norm pushes the contributions downwards. People might be temped to resolve the conflict by ignoring the feedback. In case (3), where the within-group norm is above the outgroup one, the contribution range $\overline{a}_{-i_g} < a_i < \overline{a}_{-g}$ presents another type of conflict: a within-group norm produces negative image costs and prescribes to increase the contributions, whereas the outgroup norm produced self-image benefits and prescribes to decrease contributions. In this case, people are interested in having the information on outgroup social norm.

Finally, we argue that indirect inference through $a_g(a_i)$ implies that group-level feedback is less prone to self-image costs. In particular, with *group-level* feedback the information signal works through the function $a_g(a_i)$, so that acquiring information about group-level social comparison does not impose a need to increase contribution through a self-image threat. Therefore, we can expect more ignorance of the individual vs. group-level information about the outgroup contribution norms.

We summarize our expectations with respect to information avoidance below:

Hypothesis 3. *Above-the-outgroup average contributors in relatively low cooperative groups tend to ignore the information about outgroup social norms*

Hypothesis 4. *Below-the-outgroup average contributors in relatively highly cooperative groups tend to acquire the information about outgroup social norms*

What is the effect of information avoidance/acquisition on contributions? Ignoring the information about outgroup contribution norms has a positive effect on contributions in case (3), for the contribution range $\overline{a}_{-g} < a_i < \overline{a}_{-i_g}$, where learning about being above the norm dampens the desire to improve towards the within-group average. Ignoring the information about outgroup contribution norm has negative effect on contributions in case (2), for the contribution range $\overline{a}_{-i_g} < a_i < \overline{a}_{-g}$, where learning about falling behind the norm would otherwise push the contributions upwards.

Hypothesis 5. *Ignoring feedback by below-the-outgroup average contributors in relatively highly cooperative groups has a positive effect on contributions*

Hypothesis 6. *Ignoring feedback by above-the-outgroup average contributors in relatively low cooperative groups has a negative effect on contributions*

3 Experimental procedure

This section describes our standard repeated public good game and how we introduce individual and group-level feedback on contributions. The experimental script is provided in Appendix A, and decision screens are reproduced in Appendix B.

Participants invited to a given session are randomly matched in groups of $n = 4$ and stay within the same group throughout the experiment. In each round, participants receive an endowment $w = 60$ tokens, where 1 token = CHF 0.25, and play a linear public good game with the following pay-off function:

$$\pi_i = 60 - a_i + \delta \cdot \frac{\sum_{i_g} a_{i_g}}{n}, \tag{2}$$

where $\delta = 1.6$ is a multiplier. Participants decide how many out of 60 tokens to allocate to private and group accounts.[8] Because marginal per capita return $\delta/n = 0.4 < 1$, it is individually optimal to contribute $a_i = 0$. However, since the multiplier (or marginal social benefit) $\delta > 1$, the socially optimal equilibrium is full contributions $a_i = 60$ for all $i \in i_g$. The public good game continues for 10 rounds, and one round is randomly chosen for payment.[9] All participants go through a set of control questions before the game starts.

[8] Technically, they decide how much to contribute to the group account, and the remaining amount is automatically put into their private account.

[9] Each experimental session also included a second part consisting in a simple donation task unrelated to this study. Participants are informed about the two parts of the experiment, but they receive the instructions for the second part only after the first part is finished. They also know that one of the two parts will be randomly chosen for payment. See Appendix A for exact wording.

Table 1: Overview of norm information provision across experimental conditions

	Within-group norm N_w	Individual compliance with outgroup norm N_o	Group compliance with outgroup norm N_g
Control	Yes, via payoff feedback	No	No
INDIVIDUAL FEEDBACK	Yes, via payoff feedback	Yes	No
INDIVIDUAL CHOOSE	Yes, via payoff feedback	Choice: Yes/No	No
GROUP FEEDBACK	Yes, via payoff feedback	No	Yes
GROUP CHOOSE	Yes, via payoff feedback	No	Choice: Yes/No

We specify five experimental conditions that differ in terms of the information that is provided individually after each round of the game. Table 1 presents an overview of experimental conditions. Instructions are similar for all treatments, except for text referring to alternative feedback information on the outgroup contribution norm (see Appendix A).

Starting with the control treatment, participants simply learn their earnings after every round, with no further information provided. Given their own contribution and the payoff information they can infer the within-group average contribution $\overline{a}_{-i_g}$. In the other four treatments, after every round participants also learn their earnings, and in addition they have access to information about contributions by outgroup subjects $\overline{a}_{-g}$ in that same round. With these four treatments we systematically vary both the basis for comparison (a_i or $a_g(a_i)$) and whether obtaining the feedback information is "compulsory" or results from an active choice by subjects. We now provide further details on each of these four treatments.

In the INDIVIDUAL FEEDBACK treatment, at the end of each round participants receive information about how their individual contribution compares relative to average contribution of all participants from other groups attending the same session, i.e. outgroup contribution norm $\overline{a}_{-g}$. More specifically, for each participant we compute the average contribution of all participants outside their own group in that same round and session, and provide on-screen information about whether their contribution is *below*, *equal to*, or *above* the average (i.e. their social comparison).[10] As noted above, excluding own-group members from the calculation of average ensures that the behavior of others (the outgroup norm) is payoff irrelevant information.

In the INDIVIDUAL CHOOSE treatment, we endogenize the provision of individual-level comparison feedback. Participants are informed about the very same social comparison procedure applied in the INDIVIDUAL FEEDBACK treatment, but instead of seeing the outcome of the comparison on their screen at the end of each round, they are first given additional options 'show comparison' or 'do not show comparison' (see Appendix B). Participants have to confirm their choice by clicking on their chosen option. This procedure sets no default option and thus targets lower level of ignorance in comparison to settings

[10] Note that in all treatments, we provide only relative information and no exact distance to the reference contribution. See Appendix B for screenshots.

where uncertainty is normally set as a default (Dana et al., 2007; Matthey and Regner, 2011; Grossman, 2014). If the respondent chooses to see the comparison, the exact same information as in the INDIVIDUAL FEEDBACK is shown on the screen preceding the information about individual payoff. If he chooses not to see the comparison, only individual payoff is shown. To avoid informational asymmetries within each group, this choice is private. Moreover, participants are explicitly informed that their choice to reveal relative standings to the outgroup contribution norm does not affect their payoff.

In the GROUP FEEDBACK treatment, participants learn how the total contribution within their group $a_g(a_i)$ compares with average total contributions in other groups in that session. More precisely, they are informed about whether the total contribution in their group is *below*, *equal to* or *above* the average total contributions in other groups in the room in that period. Because the comparison is also carried out with all outgroup participants in the session (i.e. $\overline{a}_{-g}$), the outgroup contribution norm is the same as in the individual feedback treatments.

In the final treatment, labeled GROUP CHOOSE, we again endogenize feedback, but this time using group-level comparison. As in the INDIVIDUAL CHOOSE treatment, participants can individually choose if they want to obtain the standing of their group relative to other groups by clicking on either 'show comparison' and 'do not show comparison' options. To avoid potential influences of this choice on within-group interaction, we do not reveal which group members choose to know the comparison. The wording is the same in individual and group-level feedback, see instructions in Appendix A and screenshots in Appendix B for further details.

Please note that since we do not provide the information on the exact value of the outgroup average contribution, participants with individual feedback cannot always infer norm compliance of their group. When above-the-average within-group contributors learn that one's individual contribution is above the average individual contribution outside the group, it does not imply that total contributions of her group is above the average total contribution is other groups. Similarly, participants with the group-level feedback cannot always infer their individual norm compliance. For the below-the-average within-group contributor, for example, learning that the total contribution of her group is above the average total contribution in other groups does not necessarily imply that her individual contribution is above the individual outgroup average.

Table 2 summarizes all possible feedback combinations on the within- and the outgroup contribution norms and provides the cross-inferences between individual vs. group comparisons to the outgroup contribution norm the participants can make.

As Table 2 shows, the conflicting feedback, "below, above" or "above, below", enables cross-inferences from individual to group comparison and prevents cross-inferences from group to individual comparison with the outgroup contribution norms. The confirming feedback, "below, below" or "above, above", in

Table 2: Inferences about outgroup contribution norms

Treatment	Comparison to within-group average	Comparison to outgroup average	Cross-inference about outgroup norm
INDIVIDUAL	below	below	No, standing of own group is unknown
	below	above	Yes, group is above average
	above	below	Yes, group is below average
	above	above	No, standing of own group is unknown
GROUP	below	below	Yes, individually above average
	below	above	No, individual standing is unknown
	above	below	No, individual standing is unknown
	above	above	Yes, individually below average

Notes: For participants whose individual or group contributions are *exactly equal* to the within-group or outgroup average, it is always possible to cross-infer their group and individual comparisons to the outgroup and the within-group norm respectively.

turn, does not allow inferring a group comparison from the individual one, but allows for inferences about the individual comparison from the group-level feedback.

We conclude this section by emphasizing that all differences between treatments concern only the inferences that can be made about individual vs. group comparison to the outgroup average contribution, i.e payoff irrelevant information.

4 Results

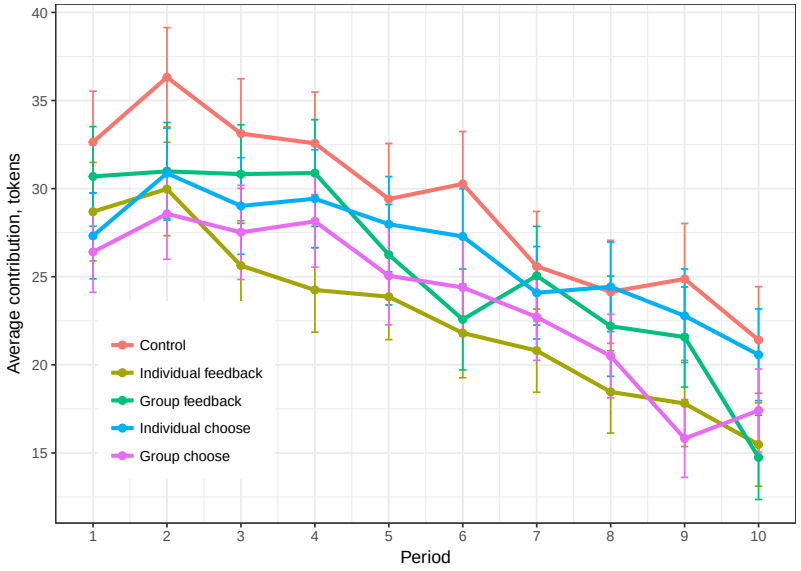
We ran 16 experimental sessions in the laboratory of the University of Neuchâtel in March 2018. Table 3 presents the descriptive summary of our sample. In total 272 participants recruited via a university mailing list took part in the experiment. The experiment is programmed in zTree (Fischbacher, 2007). Each session lasted about one hour, and the average payment was CHF 17.16.

Table 3: Summary statistics of treatments and experimental sessions

Treatment	Sessions	Participants	% females	Average age
Control	3	56	58.9	23.4
INDIVIDUAL FEEDBACK	4	52	61.5	23.8
INDIVIDUAL CHOOSE	3	56	60.7	22.4
GROUP FEEDBACK	3	52	59.6	21.8
GROUP CHOOSE	3	56	64.3	21.9

In the following, we start by reporting the aggregated treatment effects on the contributions. We then proceed to the analysis of the interaction between within- and outgroup contribution norms, and conclude with the discussion of the feedback avoidance/acquisition behavior.

Fig. 3: Average contributions by treatment (standard errors in bars)



4.1 Public good contributions

Figure 3 depicts average contributions across periods and experimental conditions. In all conditions, average contributions start at around half of the endowment (30 tokens) and gradually fall to around one third of the endowment (20 tokens) towards the end of the game. The decreasing pattern of contributions reflects a typical cooperation break-down found in public goods literature (Ledyard, 1995; Chaudhuri, 2011). Our data further suggests that contributions are highest in the CONTROL treatment, whereas INDIVIDUAL FEEDBACK performs worst, although differences are small.

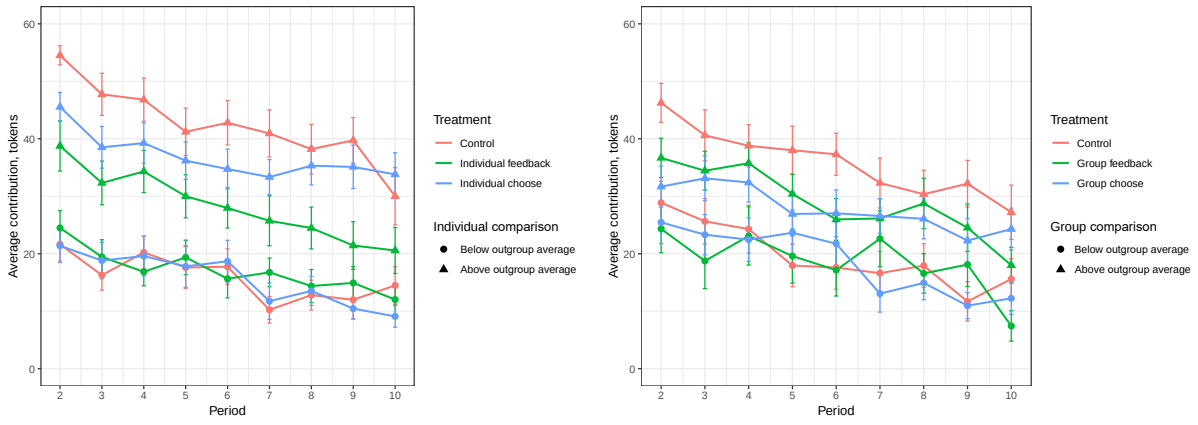
To provide further evidence about the impact of feedback information on contributions, Figure 4 reports average contribution trajectories for respondents in individual-level feedback treatments (Panel a) or group-level feedback treatments (Panel b). We condition trajectories on whether subjects have contributed above the outgroup average (positive feedback, with expected negative impact on contributions) or below the outgroup average (negative feedback, with expected positive impact on contributions).[11] We also report trajectories for the control condition as a reference, where subjects are grouped either in reference to individual (Panel a) or group-level norm compliance (Panel b).[12]

Starting with individual feedback treatments, we find that trajectories for subjects above outgroup average (upper part of Figure 4, Panel a) respond to feedback information with significantly smaller

[11] Information feedback refers to contributions in the previous period. Also, as mentioned before, it is possible that subjects are exactly on the average, which is the case for 1 choice occasion in INDIVIDUAL FEEDBACK treatment and 1 choice occasion in GROUP FEEDBACK treatment. For simplicity, these observations (not the participants) are excluded from the analysis.

[12] Note that contribution trajectories for the control condition differs in Panel a and b. The reason is that the basis for average contribution comparisons (individual or group-level) are different in each panel.

Fig. 4: Average contribution by treatment and compliance with the outgroup contribution norm (standard errors in bars)



(a) Individual feedback treatments

(b) Group-level feedback treatments

contributions relative to the control. This finding is consistent with a boomerang effect documented in Schultz et al. (2007), and can at least partially account for the fact that average contributions are lowest in the INDIVIDUAL FEEDBACK condition. This negative effect of social comparison is also present for group-level feedback, although the magnitude is smaller. This can be explained by the fact that group-level feedback information only includes individual contributions indirectly through within-group average, so that incentives to reduce own contribution is lower.

Negative (below outgroup average) feedback does not promote contributions. Indeed, trajectories for both individual and group-level feedback reported in the bottom part of Figure 4, Panels (a) and (b), are very similar to the control condition. One possible interpretation of low sensitivity to feedback is the absence of injunctive norms or moral suasion, which is a typical component of most of the field studies on social norms (e.g. Allcott, 2011; Allcott and Rogers, 2014). In a laboratory study on the influence of social norms on pro-social behavior, relying exclusively on descriptive norms, Gaechter et al. (2013) does not find an effect of the feedback on the aggregated level. Another important feature of our set-up is the payoff interdependence within groups in the public good game. As highlighted in the Sections 2 and 3, providing feedback on payoffs in all the treatments ensures that individual and group-level social comparisons do not differ in terms of possible inference about one's payoff. However, information about earnings potentially reduces sensitivity to outgroup payoff-irrelevant norms.

We now use a set of OLS random effect panel regressions to test whether the impact of feedback on outgroup contribution norm is statistically significant. Results are reported in Table 4. Column (1) simply regresses individual contributions on a set of treatment dummies, in line with Figure 3. Columns (2)

Table 4: Public good contributions across treatments and feedback types

		Individual feedback		Group-level feedback	
Contributions (a_{it})	Full sample (1)	Below average (2)	Above average (3)	Below average (4)	Above average (5)
Intercept (control)	32.67***	24.55***	48.82***	29.46***	41.32***
	(2.97)	(2.32)	(2.82)	(2.87)	(2.86)
INDIVIDUAL FEEDBACK	-6.36	-0.45	-13.04***	-	-
	(3.94)	(3.51)	(4.08)		
INDIVIDUAL CHOOSE	-2.66	1.48	-7.55*	-	-
	(4.25)	(3.03)	(3.94)		
x Ignore	-	-9.31***	7.50***	-	-
		(2.79)	(1.65)		
GROUP FEEDBACK	-3.46	-	-	-1.51	-4.49
	(3.47)			(3.63)	(3.04)
GROUP CHOOSE	-5.38	-	-	-3.50	-6.87*
	(3.75)			(3.27)	(3.85)
x Ignore	-	-	-	-2.89	-0.46
				(7.75)	(4.51)
Observations	2720	778	697	660	812
R^2	0.0534	0.0767	0.160	0.0524	0.0714

Notes: OLS random effect panel regressions reported. Column 1 includes choices by all 272 participants over nine periods (the first period with no feedback is omitted). Columns 2 and 3 consider only choices by participants in control, INDIVIDUAL FEEDBACK and INDIVIDUAL CHOOSE conditions, with own contribution in previous period respectively below (column 2) and above (column 3) outgroup average. Columns 4 and 5 consider only choices by participants in control, GROUP FEEDBACK and GROUP CHOOSE conditions, with own-group contributions in previous period respectively below (column 4) and above (column 5) outgroup average. Dummies for each time period are included in all specifications (period 2 omitted). Standard errors are clustered at the group level and reported in parentheses. *, ** and *** denote statistical significance at 10%, 5% and 1% levels respectively.

and (3) focus on the impact of individual feedback compared to the control group. We further split the sample into choices that follow negative (below average, column 2) or positive (above average, column 3) individual feedback, and add an interaction term to identify contribution decisions occurring after feedback was ignored (INDIVIDUAL CHOOSE condition). Columns (4) and (5) follow the same logic but for group-level feedback treatments plus the CONTROL group. All regressions include period fixed effects (first period omitted), and we report standard errors clustered at the group level in parenthesis.[13]

Regression results in columns (2) and (4) confirm that providing negative feedback has no positive impact on contributions. Results further show that participants in INDIVIDUAL CHOOSE treatment who ignore negative feedback display on average lower contributions, although this is not the case in the GROUP CHOOSE treatment. Columns (3) and (5) focus on positive feedback, confirming that the negative effect of outgroup social norms on contributions is much larger for individual feedback relative to group-level feedback.

[13] Contributions are censored at 0 and 60, and we have also estimated the same regressions using tobit specifications. Since results are largely similar, and we are mostly interested in deriving inference about differences in averages, we stick to OLS regressions.

4.2 Conditional cooperation: interplay between within-group and outgroup contribution norms

Our next set of results quantifies how the feedback on outgroup contribution norms affects within-group interaction. To document this issue, we split subjects according to whether they are below or above *within-group* average, a key driver of conditional cooperation, and look at how contributions increase or decrease from one period to the next depending on the sort of feedback about the outgroup contribution norm received. This is documented in Figure 5, which reports average change in contribution from previous to current period ($a_{it} - a_{it-1}$). In all panels dashed lines provide trajectories for the control group as a reference, showing that subjects who contribute above (below) their within-group average tend to decrease (increase) their contributions over time, as conditional cooperation suggests.

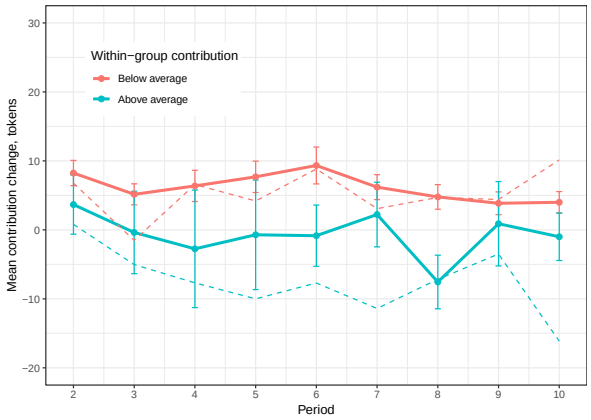
We now discuss the impact of alternative feedback on outgroup contribution norms on within-group contribution behavior. Starting with panels (a) and (b) in Figure 5, we report changes in contribution for pooled individual feedback treatments, respectively focusing on negative feedback information (below outgroup average) in Panel (a) and positive feedback information (above outgroup average) in Panel (b). We observe two main departures from the control condition. First, for subjects who contribute above their within-group average and learn that they contribute below the outgroup average (Panel a), the decline of contributions towards within-group average is much less pronounced relative to the control. In other words, if a participant learns that heor she has contributed a lot within the group, but that the contribution is small relative to outgroup average, the pace at which contribution decline is dampened. We therefore confirm the positive effect of the feedback on outgroup social norms out forward by Hypothesis 1 in Section 2.

Second, Panel (b) shows that subjects who contribute below their own-group average and receive positive individual feedback reduce their contributions compared to the previous period, whereas the within-group norm in the control suggests an increase. We therefore observe the negative effect of the feedback on outgroup social norms suggested by Hypothesis 1 in Section 2.

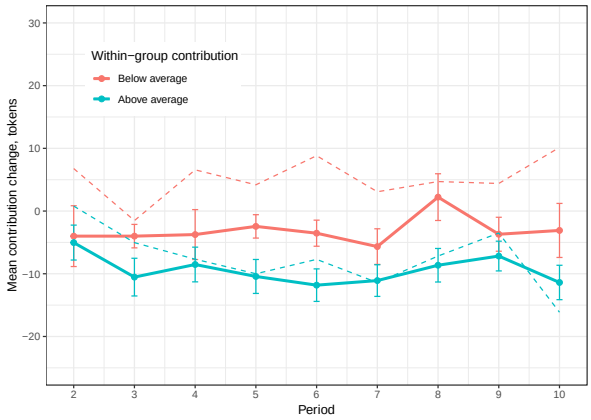
Results for group-level feedback treatments, reported in Panel (c) for below outgroup average feedback and in Panel (d) for above outgroup average feedback, show similar patterns. In particular, for subjects who contribute more than their group average and receive information that their group is below the outgroup average, the decline of contributions by these high contributors is mitigated. By contrast, for subjects who contribute below their group member average, receiving information that their group contributes more than the outgroup average dampens growth in contributions induced by conditional cooperation. The figures suggest, however, that these effect are smaller relative to those observed in the individual feedback treatments as suggested by Hypothesis 2 in the Section 2.

Corresponding regression results for changes in contributions ($a_{it} - a_{it-1}$), using OLS random effect panel models, are reported in Tables 5 and 6 for individual and group-level feedback respectively (see

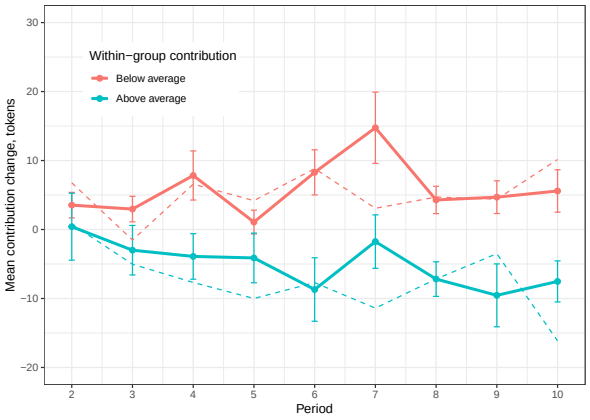
Fig. 5: Average change in contribution (control group in dashed lines, standard errors in bars)



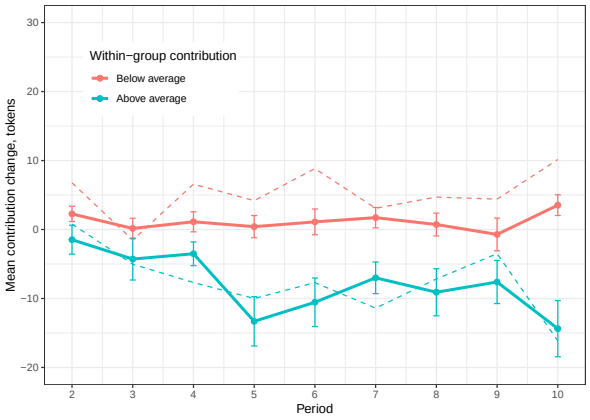
(a) Individual feedback: Below outgroup average



(b) Individual feedback: Above outgroup average



(c) Group-level feedback: Below outgroup average



(d) Group-level feedback: Above outgroup average

Huck et al., 1999; Bigoni and Suetens, 2012, for similar approaches). In both tables, the first two columns focus on exogenous social feedback conditions, while the latter two columns focus on endogenous feedback conditions. Within these, we run separate regressions for cases where contribution in the previous period was below within-group average (reported in odd columns) or above within-group average (reported in even columns). All regressions include treatment dummies, and in conditions with endogenous feedback we include an interaction term to identify instances where participants chose to ignore feedback on outgroup contribution norms. Consistent with Figure 5, all regressions include corresponding observations from the control group, so that the intercept represents conditional cooperation in the baseline. Standard-errors are clustered at the group-level.

Table 5: Within-group vs. outgroup contribution norms: individual feedback

	Control and INDIVIDUAL FEEDBACK conditions		Control and INDIVIDUAL CHOOSE conditions	
Δcontributions ($a_{it} - a_{it-1}$)	Below within-group average (1)	Above within-group average (2)	Below within-group average (3)	Above within-group average (4)
Intercept (control group)	8.19***	-8.80***	7.82***	-8.97***
	(1.92)	(1.07)	(1.80)	(1.07)
Individual feedback: Below average	-1.03	13.29***	3.28	12.69***
	(2.48)	(2.12)	(2.39)	(3.16)
x Ignore	-	-	-6.58**	-15.77***
			(2.83)	(5.74)
Individual feedback: Above average	-11.31***	-4.65**	-11.98***	-1.57
	(2.23)	(2.36)	(3.15)	(2.30)
x Ignore	-	-	5.96	3.42
			(8.56)	(2.46)
Observations	506	459	502	660
R^2	0.0291	0.0573	0.0480	0.0524

Notes: OLS random effect panel regressions reported. Column 1 and 2 consider only choices by participants in control and INDIVIDUAL FEEDBACK conditions, with own contribution in previous period respectively below (column 1) and above (column 2) within-group average. Columns 3 and 4 consider only participants in control and INDIVIDUAL CHOOSE conditions, with own contribution in previous period respectively below (column 3) and above (column 4) within-group average. Standard errors are clustered at the group level and reported in parentheses. *, ** and *** denote statistical significance at 10%, 5% and 1% levels respectively.

Table 6: Within-group vs. outgroup contribution norm: group-level feedback

	Control and GROUP FEEDBACK conditions		Control and GROUP CHOOSE conditions	
Δcontributions ($a_{it} - a_{it-1}$)	Below within-group average (1)	Above within-group average (2)	Below within-group average (3)	Above within-group average (4)
Intercept (control group)	8.11***	-8.59***	8.00***	-9.04***
	(1.89)	(1.07)	(1.86)	(1.08)
Group-level feedback: Below average	-0.70	4.24*	-1.24	3.12
	(2.67)	(2.36)	(2.31)	(2.67)
x Ignore	-	-	-3.72***	4.55
			(1.21)	(7.56)
Group-level feedback: Above average	-5.13**	-1.57	-5.82***	-0.24
	(2.39)	(1.72)	(2.01)	(2.19)
x Ignore	-	-	-0.40	-0.44
			(1.71)	(5.00)
Observations	490	470	520	477
R^2	0.0170	0.00573	0.0195	0.00168

Notes: OLS random effect panel regressions reported. Column 1 and 2 consider only choices by participants in control and GROUP FEEDBACK conditions, with own contribution in previous period respectively below (column 1) and above (column 2) within-group average. Columns 3 and 4 consider only participants in control and GROUP CHOOSE conditions, with own contribution in previous period respectively below (column 3) and above (column 4) within-group average. Standard errors are clustered at the group level and reported in parentheses. *, ** and *** denote statistical significance at 10%, 5% and 1% levels respectively.

Starting with conditional cooperation in the control group, estimates for the constant in all odd columns are positive, confirming that individuals who contribute below their within-group average increase contribution by around 8 tokens on average in the next period. Conversely, intercepts in the even columns are all negative, suggesting that participants in the control group who are above their within-group average reduce their contributions by almost 9 tokens on average.

In relation to this pattern, results in Table 5 largely consolidate the above discussion. For individuals below within-group average (columns 1 and 3) estimates associated with positive (above average) individual feedback in respect to outgroup contribution norm, $\overline{a}_{-g} < a_i < \overline{a}_{-i_g}$ are negative and highly statistically significant. This negative effect of feedback information on within-group conditional cooperation is also present for participants who contribute above within-group average, as positive feedback leads to a sharper decline in contributions relative to control (see column 2). For individuals who contribute above their within-group average (columns 2 and 4), negative (below average) individual feedback in re-

spect to outgroup contribution norm, $\overline{a}_{-i_g} < a_i < \overline{a}_{-g}$ is associated with a positive and highly statistically significant coefficient.

The possibility to ignore feedback also produces interesting findings. On the one hand, coefficients for respondents in INDIVIDUAL CHOOSE who consult feedback information are strikingly similar to those in INDIVIDUAL FEEDBACK regressions. On the other hand, those who ignore negative (below outgroup average) individual feedback display a more pronounced tendency to reduce contributions. These respondents therefore hasten the breakdown of cooperation, especially if they contribute above within-group average (column 4). By contrast, respondents who ignore positive (above outgroup average) individual feedback do not significantly differ from those who considered the feedback. In respect to our discussion of the effect of ignorance in Section 2, the negative effect defined for $\overline{a}_{-g} < a_i < \overline{a}_{-i_g}$ dominates, Hypothesis 5 is corroborated but Hypothesis 6 is not.

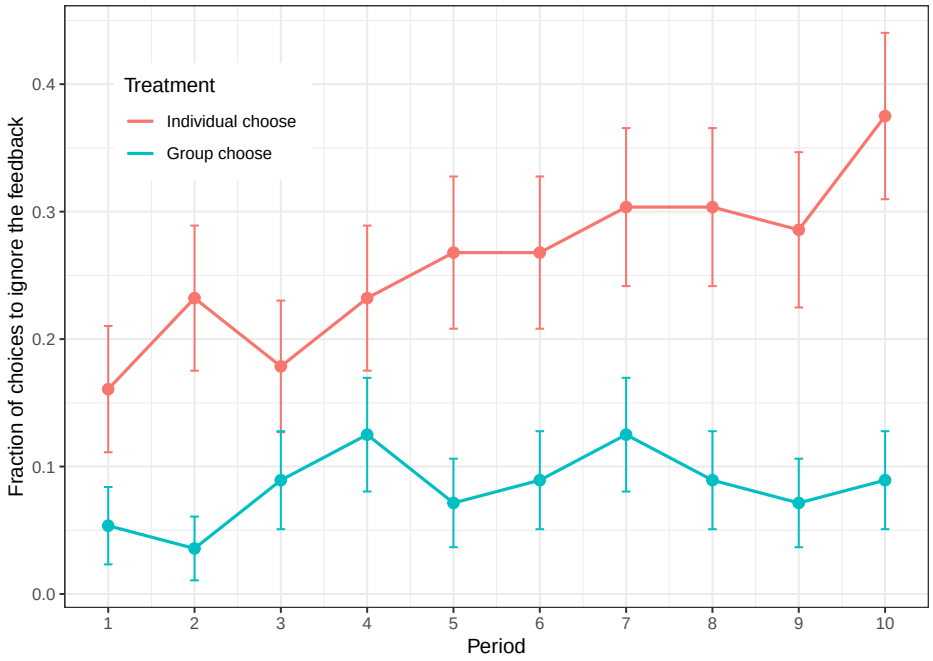
Regression results in Table 6 confirm that group-level feedback also tends to offset both positive and negative effects of conditional cooperation, although the magnitude of estimates is lower relative to individual feedback treatments. One striking difference is that those who contribute above their own-group average and ignore negative feedback do not reduce their contributions faster (Table 6, column 4).

Taken together, two conclusions about the interplay of within- and outgroup contribution norms can be made. First, social comparison feedback affects conditional cooperation through two channels: (i) negative feedback on outgroup contribution norm mitigates the breakdown of within-group cooperation for the above-the-average contributors, and (ii) positive feedback on outgroup contribution norm increases the breakdown of within-group cooperation on the below-the-average contributors. Both effects are stronger with individual feedback, but are also present for group-level feedback. Second, subjects who ignore negative feedback cooperate less than those choosing to ignore a positive feedback.

4.3 Avoidance/acquisition of the feedback on outgroup contribution norms

We conclude the analysis by scrutinizing the decisions to avoid/acquire the feedback on outgroup contribution norms. Figure 6 depicts the percentage of participants choosing to ignore the feedback on outgroup contribution norms in experimental treatments INDIVIDUAL CHOOSE and GROUP CHOOSE. Recall that these are the two treatments where subjects choose, after each round, whether or not they want to receive feedback with the comparison of their contribution to the outgroup contribution norm. On average across all 10 rounds subjects in INDIVIDUAL CHOOSE condition are around three times more likely to actively ignore feedback as compared to those in the GROUP CHOOSE condition. Specifically, subjects in INDIVIDUAL CHOOSE deliberately ignore outgroup norms in 26.1% of choice occasions, the corresponding number being 8.4% in GROUP CHOOSE.

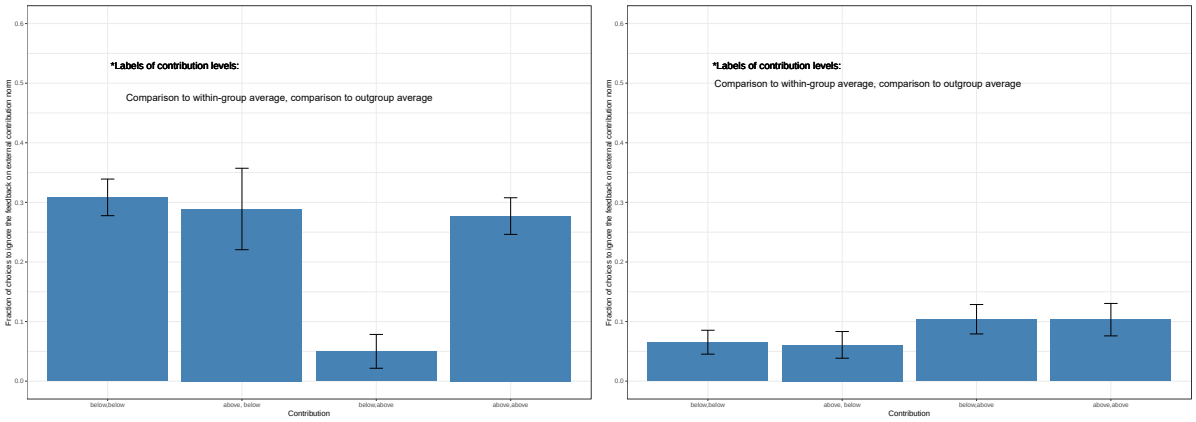
Fig. 6: Avoidance of the information about outgroup norms by treatment (standard errors in bars)



Across the 10 periods of the game, Figure 6 further shows that the difference between the two conditions is large starting from the first period of the game. After the first round, around 95% of subjects in the GROUP CHOOSE condition consult the comparison with the other groups in the session, whereas only 84% of subjects in the INDIVIDUAL CHOOSE treatment consult their individual comparison with the outgroup contribution norm (p=0.06, one-sided chi-square test). In subsequent rounds, information avoidance in the GROUP CHOOSE treatment remains low, whereas in INDIVIDUAL CHOOSE treatment information avoidance increases sharply. At the end of the 10 rounds, only 63% of subjects in the INDIVIDUAL CHOOSE treatment elect to consult the feedback, whereas the corresponding share is 91% in GROUP CHOOSE (p<0.001, one-sided chi-square test).

We next investigate how the interplay between within-group and outgroup contribution norms affect information avoidance, as put forward in the theory Section 2. In Figure 7 we again report the proportion of subjects who choose to ignore feedback on outgroup contribution norms pooled across all rounds in INDIVIDUAL CHOOSE (Panel a) and GROUP CHOOSE (Panel b). We split the data into four different cases: 1) contribution is below both within-group and outgroup average $a_i < min(\overline{a}_{-i_g}, \overline{a}_{-g})$, "below, below"; 2) contribution is above the within-group average but below the outgroup average $\overline{a}_{-i_g} < a_i < \overline{a}_{-g}$; "above, below" 3) contribution is below the within group average but above the outgroup average

Fig. 7: Avoidance of feedback on outgroup norms by treatment and contribution levels (standard errors in bars)



(a) INDIVIDUAL CHOOSE treatment

(b) GROUP CHOOSE treatment

$\overline{a}_{-g} < a_i < \overline{a}_{-i_g}$, "below, above" and 4) contribution is above both within-group and outgroup average $a_i > max(\overline{a}_{-i_g}, \overline{a}_{-g})$, "above, above"[14]

Remember from the theory section that only in cases (2) and (3) there is a conflict between within- and outgroup contribution norms. In case (2) subjects are interested in ignoring the information about the outgroup contribution norm since they are already doing better than the participants within their group, and in case (3), in contrast, people are willing to acquire information about the outgroup norm which allows them not to increase contributions towards a higher within-group average.

Two main observations can be made from Figure 7. First, the ignorance of the individual feedback is much more frequent than the ignorance of the group-level feedback on outgroup contribution norms. This finding could be interpreted as indication of higher sensitivity of self-image to individual level feedback. Second, as expected, there is a sharp difference between cases (2) and (3), for individual level feedback: when the outgroup norm pushes contributions beyond the lower within-group average it is ignored in about 30% of the time; whereas when the outgroup norm allows not to improve to the higher within-group it is ignored only in 5% of the cases. Interestingly, however, participants above the outgroup contribution norm also seem to ignore the feedback a lot. Perhaps, these participants do not want to learn about potentially free-riding in other groups and maintain a positive belief in a highly cooperative society.

[14] Note that feedback can also indicate "equal to outgroup average." In the present case, however, no subject is exactly on the outgroup average. Subjects exactly at the within-group average are excluded.

Table 7: Feedback ignorance in INDIVIDUAL CHOOSE and GROUP CHOOSE conditions

Feedback avoidance (=1)	Individual	Group
Intercept	0.32***	0.08***
	(0.07)	(0.03)
above,below	- 0.12	0.006
	(0.11)	(0.03)
below,above	-0.13*	-0.09
	(0.07)	(0.05)
above,above	-0.09	0.02
	(0.09)	(0.06)
Observations	56	56
R^2	0.012	0.00013

Notes: OLS random effect panel regressions reported. Each regression uses choices over ten periods. Dummies for each time period are included in all specifications. Standard errors are clustered at individual level and reported in parentheses. *, ** and *** denote statistical significance at 10%, 5%, 1% levels respectively.

We finish this section by reporting results from a set of random effect OLS regressions for the decisions to ignore feedback. In Table 7, the dependent variable is an 'ignorance' dummy which takes the value of one when a participant chooses to ignore the feedback and zero otherwise.

As the intercept coefficients show, participants below the within- and the outgroup average contribution, ignored the comparison in 32% of the cases in the INDIVIDUAL CHOOSE treatment and only in about 8% of the cases in the GROUP CHOOSE treatment. In INDIVIDUAL CHOOSE participants with contributions below the within-group average but above the outgroup contribution norm ignore about 13 p.p less frequently as captured by a coefficient "below, above", but the coefficient is significant only at 10% level (p=0.067). We therefore find a week support for Hypotheses 3 and 4: a significant minority of above-the-average within-group contributors ignore the negative feedback about compliance with the outgroup norm, and somewhat fewer participants from below-the-average within-group contributors ignore positive feedback about compliance with the outgroup norm.

This pattern of ignorance, however, might at least partially explain the negative effect of ignorance on changes in contributions documents in the subsection 4.2. Ignoring the disadvantageous feedback "below the outgroup average", reduces the willingness to improve, while ignoring the advantageous feedback "above the outgroup average" seems to prevent high contributors from reducing their contribution.

5 Discussion and conclusion

In this paper, we have analyzed the effect of social norms feedback in a standard public good game setting. We proposed that the effectiveness of the feedback about the norm compliance depended on

the interplay between the within-group and the outgroup contribution norms. One important distinctive feature of these two norms is pay-off relevance: within-group average contribution affects participants' pay-off, whereas outgroup average has no influence on participants' earnings.

We have shown that, when within-group and outgroup contribution norms are in conflict, outgroup norms have a stronger effect on contributions. Moreover, feedback on outgroup contribution norms both promotes and deteriorates within-group cooperation among group members, with the two effects canceling each other on aggregate. By identifying a conflict between within- and outgroup contribution norms in a public good setting, our results contribute to a growing body of literature about potential backfiring of (outgroup) social norms (Bicchieri and Dimant, 2019; Bolton et al., 2019).

We also examined the possibility to ignore the feedback on outgroup contribution norms, and observed interesting patterns. First of all, the vast majority of participants choose to receive the feedback. A significant minority of participants, however, choose not to know whether they comply with the outgroup contribution norm. In accordance with the self-image maintenance theory, people are much more likely to ignore individual vs. group level feedback. Moreover, they ignore the information on those norms that would otherwise push them to increase contributions and happily acquire the feedback on outgroup norms that suggest decrease of contributions against the prescription of the within-group cooperation. The selectivity of information acquisition through information (in)attention strategies may significantly reduce the effectiveness of social comparisons interventions.

By contrast, high within-group contributors may also ignore the information about their superior contributions compared to an outgroup contribution norm. Such ignorance could function as a self-disciplining device (Hertwig and Engel, 2016) helping high contributors avoid the temptation of following norms that would damage the well-being of their group. A closer look at both positive and negative effects of information acquisition strategies in a broader context of social norms promises an intriguing venue for future research.

Appendix A Experimental script

These instruction were printed and handed out to participants. The text is the same for all treatments, except for the part that discusses information feedback after all participants have made their contribution decision.

Instructions

Welcome and thank you for participating in this experiment. Please switch off your mobile phones and do not talk to other participants. If you have a question, raise your hand, and the experimenter will answer you privately.

In this experiment you can earn money. Your earnings depend on your own decisions, the decisions of other participants as well as on chance.

During the experiment your earnings will be expressed in Tokens which will be converted at the end of the experiment in CHF at the following conversion rate:

1 Token = 0.25 CHF.

During the whole experiment your anonymity is warranted.

Experimental procedure

The experiment consists of two parts. The following instructions explain the first part of the experiment. The instructions for the second part of the experiment will be presented to you after you have finished the first part.

After the second part of the experiment is over, the computer will randomly select either first or the second part of the experiment to define your final payment.

It means that you will either receive the payment from part 1 or from part 2 of the experiment. Since both parts can be selected with equal probability, please take all the decisions carefully in both parts of the experiment.

Part 1

At the beginning of this part of the experiment, the participants will be matched in the groups of four. The participants remain in the same groups throughout all the 10 rounds in this part of the experiment.

At the beginning of every round, each participant receives an endowment of **60 Tokens**. Next, the participants will be asked to make a decision regarding this endowment.

In each group every participant has its own **account A**, and the group has a common **account B**. It means, in every group there are four private accounts A, one for each participant, and a single common account B.

The participants decide how they would like to allocate their endowment of 60 tokens between accounts A and B.

For each token the participants put on their private account A, they receive exactly one token. The tokens the participants in a group put on the group account B are summed up and **multiplied by 1.6**. The resulting total amount of tokens on the account B is then divided equally among the four members of the group.

The payment of each participant is thus defined as follows:

tokens on account A + **0.4***sum of contributions to group account B

In means that your earnings from account A are independent from the decisions of other participants. You simply keep the money that you put on account A.

Your payment from the account B, however, depends on the decisions of all group members.

It holds that you always receive 0.4 of the sum of the contributions to the group account B, regardless of how many tokens you contribute to this account; the more tokens are contributed to the group account B, the higher are the total earnings of the group (since the value of each contribution is increased by 60%).

Consider the following example:

Group member	1	2	3	4
Endowment	60	60	60	60
Contribution to group account B	0	20	40	60
Sum of contributions	120			
Profit from group account	120*1.6/4=48			
Tokens kept on private account A	60	40	20	0
Profit for round	108	88	68	48

After all the group members have made their decisions, you will be informed about your payment in that round.

Individual feedback:

Every round once all the participants have made their decisions, the computer will compare your contribution to the group account B with the average contribution of all the other participants (your group-members are not included in this comparison). Your relative standing is then defined as below, above or equal to the average. After the comparison is made, your relative standing will be shown to you on the screen. The information about your relative standing does not affect your payment.

Individual choose:

Every round once all the participants have made their decisions, you will have the opportunity to compare your contribution to the group account B with the average contribution of all the other participants (your group-members are not included in this comparison). Your relative standing is then defined as below, above or equal to the average. If you choose to make a comparison, your relative standing will be shown to you on the screen. It means that you decide individually whether to know your relative standing. Your decision to display or not to display your relative standing will not be disclosed to other participants. The information about your relative standing does not affect your payment.

Group feedback:

Every round once all the participants have made their decisions, the computer will make a comparison of the total contribution of your group to the group account B with the average total contributions to the account B of the other groups. The relative standing of your group is then defined as below, above or equal to the average. After the comparison is made, the relative standing of your group will be shown to you on the screen. The information about the relative standing of your group does not affect your payment.

Group choose:

Every round once all the participants have made their decisions, you will have the opportunity to compare the total contribution to a group account B of your group to the average total contribution to account B of the other groups. The relative position of your group is then defined as below, above or equal to the average. If you choose to make a comparison, the relative standing of your group will be shown to you on the screen. It means that you decide individually whether to know the relative standing of your group. Your decision to display or not to display it will not be disclosed to other participants. The information about the relative standing of your group does not affect your payment.

You will then be informed about your payment in that round.

Once the 10 rounds are over, the computer will randomly select one round to define your payment for this part of the experiment. It means that your payment in this part of the experiment will be determined according to the decisions made in that particular round based on the aforementioned procedure. Since every round can be selected with equal probability, please take all the decisions carefully.

The experiment will then proceed to the second part. The experiment will end with the questionnaire which does not affect your payment. Should you read the instructions carefully, please click on “continue”. Before the experiment begins, you will be asked a few control questions to test your understanding of the experimental procedure.

Good Luck.

Appendix B Decision screens

Fig. B1: Contribution stage

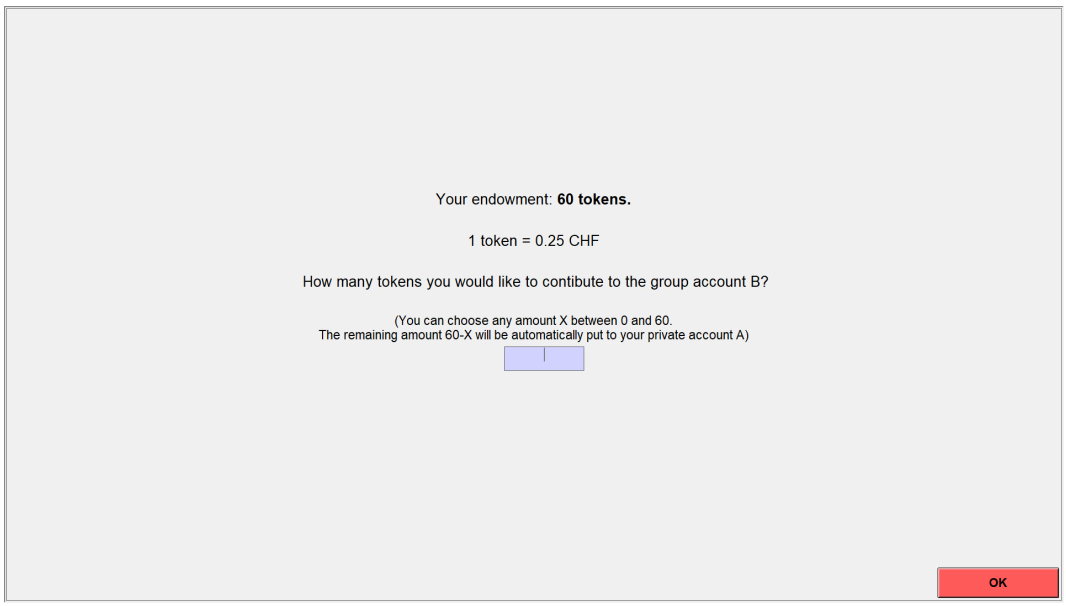


Fig. B2: Individual choose: choice to ignore the feedback

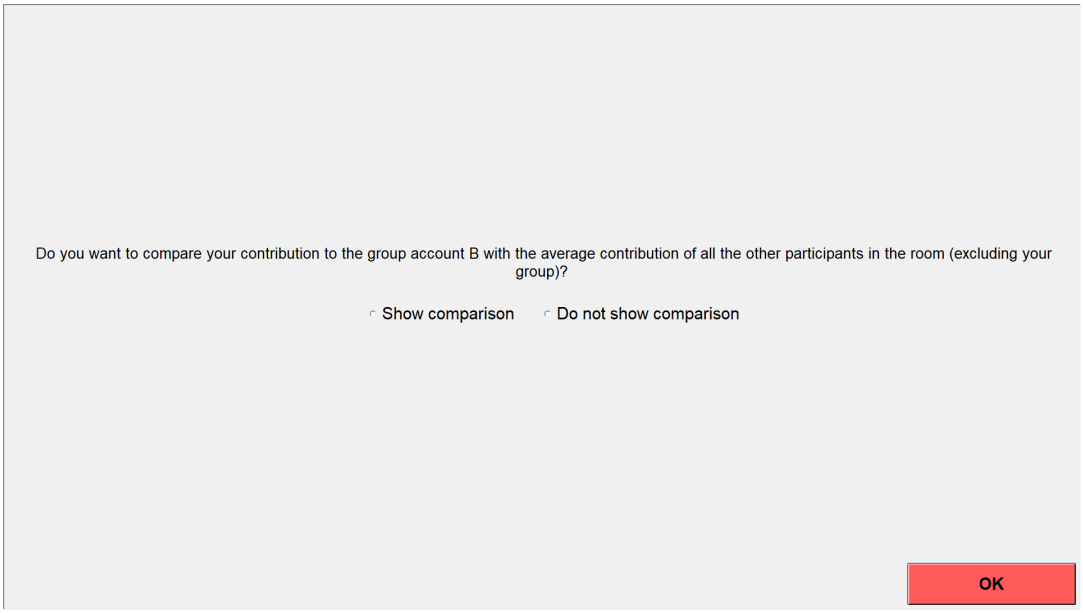


Fig. B3: Individual feedback: both individual feedback treatments

Fig. B4: Group choose: choice to ignore group-level feedback

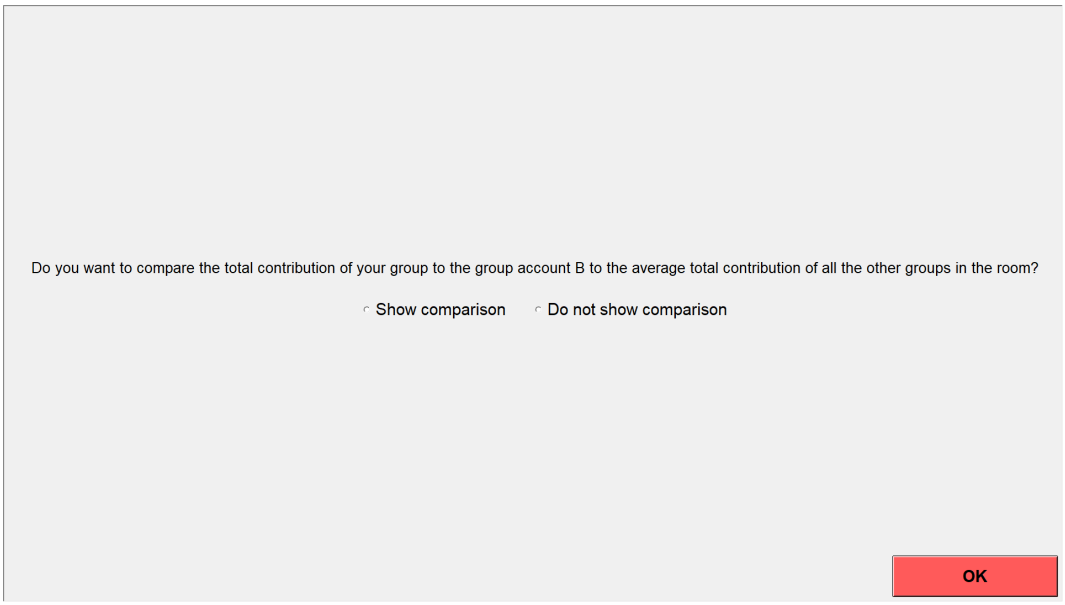
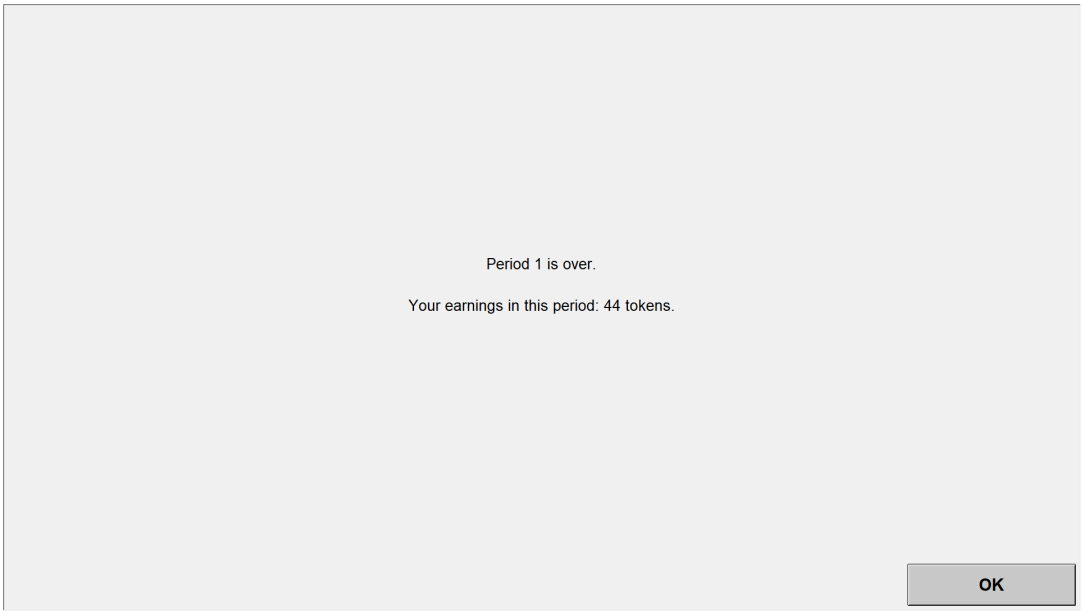


Fig. B5: Group feedback: both group feedback treatments

Fig. B6: Payoff feedback: all treatments, no choice.



References

Allcott H (2011) Social norms and energy conservation. Journal of Public Economics 95(9):1082 – 1095

Allcott H, Kessler J (2015) The welfare effect of nudges: A case study of energy use social comparisons, nBER Working Paper Series

Allcott H, Rogers T (2014) The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. American Economic Review 104(10):3003–37, DOI 10.1257/aer.104.10.3003, URL http://www.aeaweb.org/articles?id=10.1257/aer.104.10.3003

Ayers U, Raseman S, Shih A (2013) Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. Journal of Law, Economics and Organization 29:992–1022

Babcock P, Bedard K, Charness G, Hartman J, Heather R (2015) Lettind down the team? Social effects of team incentives. Journal of the European Economic Association 13(5):841–870

Bandiera O, Barankay I, Rasul I (2012) Team incentives: Evidence from a firm level experiment. Journal of the European Economic Association 11

Bicchieri C, Dimant E (2019) Nudging with Care: The Risks and Benefits of Social Information. Working Paper April 2019, Behavioral Ethics Labs, University of Pennsylvania

Bicchieri C, Xiao E (2009) Do the right thing: but only if others do so. Journal of Behavioral Decision Making 22(2):191–208

Bigoni M, Suetens S (2012) Feedback and dynamics in public good experiments. Journal of Economic Behavior & Organization 82(1):86 – 95, DOI https://doi.org/10.1016/j.jebo.2011.12.013, URL http://www.sciencedirect.com/science/article/pii/S0167268111003143

Boehm R, Rockenbach B (2013) The inter-group comparison - intra-group cooperation hypothesis: Comparisons between groups increase efficiency in public goods provision. PLOS ONE 8:1–7

Bolton G, Dimant E, Schmidt U (2019) When a Nudge Backfires. Using Observation with Social and Economic Incentives to Promote Pro-Social Behavior. PPE Working Papers 2019-03, Center for Decision Research and Expeirmental Economics, University of Nottingham

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

Costa DL, Kahn ME (2013) Energy conservation "nudges" and environemtalist ideology: Evidence from a randomized residential electricity field experiment. Journal of the European Economic Association 11(3):680–702

Dana J, Weber R, Kuang J (2007) Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. Economic Theory 33(1):67–80

Eckel CC, Fatas E, Godoy S, Wilson RK (2016) Group level selection increases cooperation in the public good game. PLOS ONE 11:1–13

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

Engel C, Kube S, Kurschilgen M (2011) Can we manage first impressions in cooperation problems? An experiment. Discussion Paper Series of the Max Planck Institute for Research on Collective Goods 2011_05, Max Planck Institute for Research on Collective Goods

Ferraro PJ, Miranda JJ, Price MK (2011) The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. American Economic Review 101(3):318–22

Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2):171–178

Fischbacher U, Gachter S (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. American Economic Review 100(1):541–56, DOI 10.1257/aer.100.1.541, URL http://www.aeaweb.org/articles?id=10.1257/aer.100.1.541

Fischbacher U, Gaechter S, Fehr E (2001) Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71(3):397 – 404, DOI https://doi.org/10.1016/S0165-1765(01)00394-9, URL http://www.sciencedirect.com/science/article/pii/S0165176501003949

Gaechter S, Nosenzo D, Sefton M (2013) Peer effects in pro-social behavior: Social norms or social preferences? Journal of the European Economic Association 11(3):548–573, DOI 10.1111/jeea.12015, URL +http://dx.doi.org/10.1111/jeea.12015

Gaechter S, Gerhards L, Nosenzo D (2017) The importance of peers for compliance with norms of fair sharing. European Economic Review 97:72 – 86

Grossman Z (2014) Strategic ignorance and the robustness of social preferences. Management Science 60(11):2659–2665

Grossman Z, van der Weele JJ (2017) Self-image and willful ignorance in social decisions. Journal of the European Economic Assosiation 15:173–217

Hallsworth M, List JA, Metcalfe RD, Vlaev I (2017) The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. Journal of Public Economics 148:14 – 31

Harding M, Rapson D (2017) Do voluntary carbon offsets induce energy rebound? A conservationist's dilemma, working paper

Hertwig R, Engel C (2016) Homo ignorans: Deliberately choosing not to know. Perspectives on Psychological Science 11(3):359–372

Ho B, Taber J, Poe G, Bento A (2016) The effects of moral licensing and moral cleansing in contingent valuation and laboratory experiments on the demand to reduce externalities. Environmental & Resource Economics 64(2):317–340

Huck S, Normann H, Oechssler J (1999) Learning in cournot oligopoly: An experiment. The Economic Journal 109 (454):C80–C95

Ito K, Ida T, Tanaka M (2018) Moral suasion and economic incentives: Field experimental evidence from energy demand. American Economic Journal: Economic Policy 10(1):240–67, DOI 10.1257/pol.20160093, URL http://www.aeaweb.org/articles?id=10.1257/pol.20160093

Jacobsen GD, Kotchen MJ, Vandenbergh MP (2012) The Behavioral Response to Voluntary Provision of an Environmental Public Good: Evidence from Residential Electricity Demand. European Economic Review 56:946–960

Karlin B, Zinger JF, Ford R (2015) The effects of feedback on energy conservation: A meta-analysis. Psychological Bulletin 144:1205–1227

Kluger AN, DeNisi A (1996) The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin 199:254–284

Krupka E, Weber R (2009) The focusing and informational effects of norms on pro-social behavior. Journal of Economic Psychology 30 (3):307–320

Ladley D, Wilkinson I, Young L (2015) The impact of individual versus group rewards on work group performance and cooperation: A computational social science approach. Journal of Business Research 68(11):2412 – 2425, DOI https://doi.org/10.1016/j.jbusres.2015.02.020, URL http://www.sciencedirect.com/science/article/pii/S0148296315001022

Ledyard JO (1995) Public Goods: A Survey of Experimental Research. In: Kagel J, Roth A (eds) Handbook of experimental economics, Princeton University Press, chap 2

Lickel B, Schmader T, Curtis M, Scarnier M, Ames DR (2005) Vicarious shame and guilt. Group Processes & Intergroup Relations 8(2):145–157

Matthey A, Regner T (2011) Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior. Games 2(1):114–135

Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: A theory of self- concept maintance. Journal of Marketing Research 45(6):633–644

Robinson E, Thomas J, Aveyard P, Higgs S (2014) What everyone else is eating: A systematic review and meta-analysis of the effect of informational eating norms on eating behavior. Journal of the Academy of Nutrition and Dietetics 114(3):414 – 429, DOI https://doi.org/10.1016/j.jand.2013.11.009, URL http://www.sciencedirect.com/science/article/pii/S2212267213016808

Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. Psychological Science 18(5):429–434

Sutter M, Czermak S, Feri F (2013) Strategic sophistication of individuals and teams. experimental evidence. European Economic Review 64:395 – 410

Tajfel H, Turner JC (1986) The social identity theory of intergroup behaviour. In: Worchel S, Austin WG (eds) Psychology of Intergroup Relations, 2nd edn, Nelson"=Hall Publishers, Chicago, pp 7–24

Tiefenbeck V, Staake T, Ross K, Sachs O (2013) For Better or for Worse? Empirical Evidence of Moral Licensing in a Behavioral Energy Conservation Campaign. Energy Policy 57:160–171

Welten SCM, Zeelenberg M, Breugelmans SM (2012) Vicarious shame. Cognition and Emotion 26(5):836–846