Diederich, Johannes; Goeschl, Timo; Waichman, Israel

Working Paper

Self-nudging is more ethical, but less efficient than social nudging

AWI Discussion Paper Series, No. 726

Provided in Cooperation with:
Alfred Weber Institute, Department of Economics, University of Heidelberg

Suggested Citation: Diederich, Johannes; Goeschl, Timo; Waichman, Israel (2023) : Self-nudging is more ethical, but less efficient than social nudging, AWI Discussion Paper Series, No. 726, University of Heidelberg, Department of Economics, Heidelberg, https://doi.org/10.11588/heidok.00033230

This Version is available at:
https://hdl.handle.net/10419/278454

HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Self-nudging is more ethical, but less efficient than social nudging

Johannes Diederich

Timo Goeschl

Israel Waichman

AWI DISCUSSION PAPER SERIES NO. 726
April 2023

Self-nudging is more ethical, but less efficient than social nudging*

Johannes Diederich†, Timo Goeschl‡ and Israel Waichman§

This version: April 25, 2023

Abstract

Manipulating choice architectures to achieve social ends ('social nudges') raises problems of ethicality. Giving individuals control over their default choice ('self-nudges') is a possible remedy, but the trade-offs with efficiency are poorly understood. We examine under four different information structures how subjects set own defaults in social dilemmas and whether outcomes differ between the self-nudge and two exogenous defaults, a social (full cooperation) and a selfish (perfect free-riding) nudge. Subjects recruited from the general population ($n = 1{,}080$) play a ten-round, ten-day voluntary contribution mechanism online, with defaults triggered by the absence of an active contribution on the day. We find that individuals' own choice of defaults structurally differs from full cooperation, empirically affirming the ethicality problem of social nudges. Allowing for self-nudges instead of social nudges reduces efficiency at the group level, however. When individual control over nudges is non-negotiable, self-nudges need to be made public to minimize the ethicality-efficiency trade-off.

JEL Classification: H41, C92, D91

Keywords: Choice architecture; defaults; public goods; self-nudge; online experiment.

*This research was generously funded by the Innovation Fund FRONTIER at Heidelberg University (ZUK 49/2 5.2.141).

†Department of Economics, University of Kassel, Germany. Email: johannes.diederich@uni-kassel.de

‡Department of Economics, Heidelberg University, Bergheimerstrasse 20, 69115 Heidelberg, Germany. Email: goeschl@uni-heidelberg.de

§Bard College Berlin. Platanenstrasse 24, 13156 Berlin, Germany. Email: i.waichman@berlin.bard.edu

1 Introduction

'Nudging'[1] has become popular among policy-makers over the last decade (Sunstein and Reisch, 2013; Halpern, 2015). The use of nudges in behavioral public policy, however, and the choice architects behind them have also attracted criticism. It has been argued, for example, that the choice architect uses, without consent, people's inertia or inattention against them (Sunstein, 2017). There has also been concern about unintended effects of nudges when the choice architect overlooks policy-relevant heterogeneity among the nudged population (Thunström et al., 2018). One of the most common objections has been against employing nudges in settings in which the interests of the choice architect and the decision maker are not necessarily aligned (Altmann et al., 2013; Sunstein, 2015; Hagman et al., 2015). Social dilemmas are the prime example for such settings. There, "social nudges" (Nagatsu, 2015) are deployed with the intention of delivering choice outcomes that generate higher benefits not necessarily to each individual, but to society (or a group) as a whole. An individual targeted by a social nudge may therefore have reason to be distrustful of choice architects who are likely to prioritize social objectives over the individual's own benefits (Sunstein, 2015). This is largely unproblematic when the individual's interests coincide with those of the group: People approve of social nudges that support societal outcomes they favor (Tannenbaum et al., 2017). When the interests do not align, however, people disapprove of a social nudge, particularly when it is "perceived as foolish, wrong, harmful, expensive, or as the imposition of some high-minded [...] elite" (Sunstein and Reisch, 2013). Such disapproval can lead to "psychological reactance" among the 'nudged' population and threaten or even reverse the effect of the nudge (Arad and Rubinstein, 2018).

When a lack of alignment between the choice architect and the targeted individuals threatens the efficiency and ethicality of nudging, then one option is to increase alignment by personalizing the nudge on the basis of additional data about the individual (Sunstein, 2013).[2] A second option, fully consistent

[1] *Nudging* has been defined as a deliberate manipulation of the 'choice architecture', that is, the non-price elements of the economic environment in which people take decisions. Originating from the philosophical position of liberal paternalism, the goal of the manipulation is to alter people's decisions in a beneficial direction while maintaining freedom of choice across options and their associated economic incentives (Thaler and Sunstein, 2008). Because of its potential to counteract well-known biases in human decision-making, nudging promises to deliver choice outcomes that generate – for 'the nudged' – higher own benefits than those that would arise in its absence.

[2] Personalized defaults acknowledge the heterogeneity among the nudged population and can reduce the problem of one-size-fits-all present in uniform nudges. Personalized nudges are likely to require gathering large amounts of personal data to improve the fit of the nudge

with the principles of liberal paternalism, is to give individuals themselves control over the choice architecture. In Thaler and Sunstein's words, this would mean *"set[ting] the default by asking what reflective [individuals] would actually want."* (Thaler and Sunstein, 2008). Like nudging itself, the idea of such "self-nudging"[3] is not new: Users of fitness and health applications on smartphones set the frequency and timing of reminders, feedback, and other defaults (Caraban et al., 2019). People with weight management issues deliberately remove high-calorie items from their line of sight in the office and at home (Bucher et al., 2016). Members of mutual funds set their own default combination of dividend payout and reinvestment when they join an automatic dividend reinvestment plan (Feito-Ruiz et al., 2020).[4] The informational and ethical advantages of giving people control over the choice architecture that structures their environment have been highlighted in the philosophical literature (Reijula and Hertwig, 2020). Self-nudging could also have material advantages: In an experiment, subjects used the control opportunity to achieve higher own benefits by anticipating and avoiding behavioral biases (Tontrup and Sprigman, 2019). These observations point to the potential of self-nudging to help people make better decisions for their own benefit (Banerjee and John, 2021).

For social dilemmas, in which "better" decisions are intended to generate higher group benefits, self-nudging has so far received little attention.[5] This is despite its potential to overcome the objections against social nudges imposed exogenously that were discussed above. The present paper contributes towards closing this research gap.Specifically, it reports on an online experiment that compares – in a paradigmatic social dilemma – the performance of exogenously chosen nudges with that of endogenous nudges in three dimensions: how individuals choose to set their own nudge (default choice), how self-nudges affect

(Thaler and Tucker, 2013; Yeung, 2017) such as past behavior (Briscese, 2019). The privacy dimensions of such data-intensive personalization raise ethics issues of their own. There are also concerns whether personalized nudges can be reconciled with notions of universality and equal treatment that apply to public policy (Mills, 2020).

[3]In the literature, the idea has also been discussed under the heading of "self-management" (Schelling, 1978) or "behavioral self-management"(Tontrup and Sprigman, 2019).

[4](Automatic) reinvestment plans determine the default action of a member of a mutual fund in the absence of an active reinvestment decision, but the member can deviate from the default at any time. Members can set the default to contributing all of their annual dividends back to the fund, receive all of their annual dividends as a cash pay-out, or some combination of the two.

[5]One exception are Engel and Kurschilgen (2020) who investigate the effect of normative beliefs about (minimum) contributions to the public good on actual contributions in a public goods game. They find that the normative belief elicitation increases cooperation only when the baseline cooperation is rather low. The present paper employs an altogether different approach based on participants setting their own default contributions.

individual behavior (payoffs), and how they affect group outcomes (efficiency).

The experimental design combines four components. First, as the social dilemma, it employs the standard linear public goods game, or voluntary contribution mechanism (VCM, e.g., Isaac and Walker, 1988; Fehr and Gächter, 2000; Kimbrough and Vostroknutov, 2016). The strength of using the VCM lies in the clear and cardinal metric for measuring how individuals self-nudge and how effective the outcomes are for individuals and the group.

Second, for the nudge, the design uses choice defaults, specifically default contribution levels in the VCM. Defaults are prototypical tools in the hands of the choice architect (Madrian and Shea, 2001; Thaler and Benartzi, 2004; Thaler and Sunstein, 2008) and have been implemented in economic experiments on social dilemmas before (Fosgaard and Piovesan, 2015; Bruns and Perino, 2021). Defaults have been among the most effective nudge policy intervention (e.g., Johnson and Goldstein, 2003; Madrian and Shea, 2001). Our design considers two exogenous nudges set by a choice architect, the social nudge of contributing the full endowment and the selfish nudge of contributing nothing, plus – as a key innovation – self-nudges set individually by experimental subjects.

The third component is the intervention point for the nudge. Arguably, the most natural point for a default intervention is the participation stage. This is the stage after the group has been formed, but before members take their active contribution decisions. By departing from the procedures of standard laboratory experiments, the design reflects this consideration: Subjects are assembled into groups during the first of multiple sessions and need to re-engage anew for each session. There, setting a default for the case of non-participation is a necessary feature, in particular when the multi-session experiment lasts several days (Isaac et al., 1994; Normann et al., 2014; Diederich et al., 2016). This mirrors field cases of public good provision in which contributions are determined by defaults such opt-in and opt-out options common in charitable donations, marketing, church fees, pension plan, etc.[6] This means that we consider non-participation defaults that manipulate what contribution decision will be taken on behalf of those group members that fail to show up for the contribution stage.[7]

[6]In principle, the nudge could operate at one of three stages of the VCM: Group formation (Ahn et al., 2008), group participation (Cason et al., 2004), and group contributions (Fosgaard and Piovesan, 2015). The participation stage has attracted only limited attention in economic experiments: In laboratories subjects are already seated in front of their screens by the time that they have to take decisions. In online settings, however, experimenters have been encountering the participation stage in the form of attrition problems: Subjects drop out or fail to reliably participate in every round of interaction (Arechar, 2018; Horton, 2011; Shank, 2016).

[7]It has been suggested to us that endogenous nudges are analogous to "snudges". Kaiser

The fourth component of the design is the explicit consideration of variations in feedback information. Information about defaults and contributions can each be kept private (known only to the subject) or made public (communicated to the group). Feedback information about others' past contributions has been shown to affect own contributions in public good games (e.g. Neugebauer et al., 2009; Angelovski et al., 2018), primarily through the beliefs channel. To the extent that knowing about others' defaults affects subjects' beliefs, making defaults public could therefore also affect contributions.

The repeated VCM was played online in fixed groups of four over the course of ten days, with one contribution decision to be taken each day. A total of 1,080 members of the general population in Germany participated in the experiment. There were three design factors. The first factor is which of the three different non-participation default contributions apply. These step in when a member fails to make his or her daily contribution decision. The natural baseline default is an 'exogenous default' of perfect free-riding: Non-participating group members make a zero contribution in that round. This 'selfish' nudge is the almost universal default in multi-session VCM experiments (e.g. Isaac et al., 1994; Diederich et al., 2016). A second exogenous default is of full cooperation: Non-participating members contribute their entire per-round endowment to the public good in that round. These 'social' nudges are used, for example, in the one-shot VCM experiments by Altmann and Falk (2009) and Hokamp and Weimann (2021). Against these two benchmarks, we allow for a third possibility: endogenous contribution defaults that faithfully implement Thaler and Sunstein's vision. Here, group members are asked to set individually and irrevocably their own, unique default level in the first of the ten-round interactions. In case of non-participation, members then passively contribute this personally chosen amount of the per-round endowment in that round. The second factor is whether group members are informed about the default that applies to other group members or not. The third factor is whether, after every round, the number of active participants and average contribution of the other group members are made public to the group or not. The second and third factor jointly determine under which of the four 'information structures' the VCM operates. This results in a full factorial design: Each of the three default rules for the VCM is implemented under each of the four information structures.

Our results focus on two crucial criteria for comparing the three default rules, holding the information structure constant. One is the "ethicality" di-

et al. (2020) define them as "offering self-binding commitments". However, there is no commitment component implicit in subjects setting their own defaults.

mension. This criterion relates to the question whether the default rules reflect the participants' default preferences, as revealed through self-determined nudges. The other is the "efficiency" dimension: This criterion relates to the material gains that arise for the groups under the three different default rules. The standard information structure for our results is one with private defaults and public contributions. This information structure is a natural choice: Private providers of public goods, such as charities or public-service associations, periodically disclose total contributions received, but do not disclose whether members gave through a monthly gift plan or made discretionary donations. Results for the other three information structures are then compared and contrasted to examine the generalizability under considerations of robustness and of information structure design to enhance cooperation.

Under the standard information structure, there are two main findings: First, exogenous defaults do raise problems of ethicality: When subjects could set their own defaults, they chose an average non-participation contribution of 44 percent of the endowment. Fewer than ten percent of the subjects chose either the perfect free-riding (0 percent) or full cooperation (100 percent) default. The self-nudge therefore rarely coincides with the full cooperation default chosen by a choice architect who prioritizes group benefits. This discrepancy illustrates the problem of ethicality associated with the lack of alignment between choice architect and individual and underlines the ethical advantages of self-nudging. The second finding is that there is a trade-off between *ethicality* and *efficiency*: Compared to the perfect free-riding default, total contributions under the ethically problematic full cooperation default were 30 percent higher. The ethically unproblematic self-nudges, on the other hand, did not significantly lift total contributions above those of the free-riding default, despite the higher non-participation default of 44 percent of endowment.

Extending the analysis from the standard to all four information structures, we find that most, but not all findings generalize. One finding that generalizes is that exogenous defaults are *ethically problematic*: The median self-nudge was 50 percent of endowment and less than 10 percent of subjects chose either a perfect free-riding or a full cooperation default, with no significant differences across information structures. Results for the efficiency dimension also generalize: The (ethically problematic) social nudge of full cooperation significantly raises total contributions relative to a perfect free-riding default, irrespective of the information structure. Compared to the self-nudge, however, it significantly increases total contributions only when defaults are private. When defaults are

public information, the (ethically unproblematic) self-nudge is no longer significantly less efficient than the (ethically problematic) full cooperation default. In other words, public defaults narrow the gap between social nudges and self-nudges, but social nudges under private defaults still outperform self-nudges under public defaults.

Our findings merit attention: They show that when efficiency is the main criterion, the social planner should be made the choice architect, leading to an ethically problematic one-size-fits-all social nudge. When ethicality is the main criterion, individuals should be chosen as choice architects of their own self-nudge. The trade-off between efficiency and ethicality can be mitigated by changes in the information structure, but not fully resolved. Making self-nudges public, for example, raises efficiency, but not to levels achievable by a social nudge.

Beside our conceptual contribution, our study demonstrates how to use attrition productively to inject realism into the study of defaults, status-quo bias, and participation decisions in social dilemmas. The experimental setting of an online multi-day VCM (Isaac et al., 1994; Diederich et al., 2016) captures many features of real-world interactions. There, small but positive participation costs prevent participants in the interaction from making an active decision every time a decision is to be made (Pecorino and Temimi, 2007; Osborne et al., 2000). This realism is typically absent in laboratory experiments. In online experiments, attribution is typically regarded as a nuisance factor (Arechar et al., 2018). Thus, our approach could be useful for addressing research questions in which increased realism is a step towards greater external validity.

The structure of the paper is as follows. Section 2 presents the experimental design and procedures. Section 3 presents the results for the standard information structure of private defaults and public contributions. Section 4 explains which of these results generalize to the other information structures. Section 5 presents exploratory analysis on default setting and beliefs. Finally, concluding remarks are provided in Section 6.

2 Experimental Design and Procedure

This section describes in detail the workhorse game and treatments, the online recruitment procedure, and the experimental procedure.

2.1 Design

We employ a standard VCM (e.g., Isaac and Walker, 1988; Fehr and Gächter, 2000) in groups of four and with an MPCR of 0.4. The game was repeated for ten rounds in partner-matching. In each round, subjects were endowed with 80 units of the experimental currency, which they could divide between a private account and a common group account. Each currency unit allocated to the private account would increase a subject's payoff by one unit, while each unit allocated to the group account would increase each group members' payoff by 0.4 units. Thus, the payoff, π, to an individual i in any given round is given by:

$$\pi_i = 80 - m_i + 0.4 \left(m_i + \sum_{j \neq i}^{4} m_j \right) \tag{1}$$

Equation 1 captures the social dilemma structure that (i) for a payoff-maximizing subject, there is a dominant strategy to allocate all her endowment to her private account, and (ii) the resulting outcome is Pareto-dominated by the case where all subjects allocate all their endowments to the group account.

To study the effect of (self-)nudging at the participation stage of the social dilemma, we use a "multiple session" variant of the VCM (Isaac et al., 1994; Diederich et al., 2016). In multiple session experiments, rounds typically last several days so that subjects depart from and return to the experiment for each single round. This design feature forces researchers to cope with attrition. For our purposes, attrition provides a natural way to introduce contribution default rules for the case of subjects not participating in a given round. Changing the default rule in a multiple session VCM does neither reduce or alter the choice set nor does it inherently change the economic incentives (Thaler and Sunstein, 2008). However, once introduced, making an active decision (and thus deviating from the default contribution decision) arguably incurs some small non-monetary costs of cognitive effort and time to overcome behavioral inertia. These small costs are commonly invoked to explain why defaults have the ability to "stick" (Blumenstock et al., 2018). Each invitation email to a new round reminded subjects that a non-participation default had been set.

Our experiment consists of a $3 \times 2 \times 2$ factorial design. The first factor is which of the three default rules applies, the second whether the defaults applying to the group are kept private or made public, and the third whether contributions are kept private or made public. Table 1 presents the different treatments and the number of subjects/groups per treatment.

The three default conditions. The first default condition, SELFISH, exogenously sets the default contribution to perfect free-riding. That is, a subject would automatically place all of her experimental endowment in her private account whenever she did not submit an active decision in a given round. This is the 'selfish nudge'. The second default, SOCIAL, exogenously sets the default contribution to full cooperation. That is, a subject would contribute all endowment to the common account whenever she did not submit an active decision in a given round. This is the 'social nudge'. In the third default, SELF, we asked each subject to choose individually and irrevocably a default contribution amount for herself after her first round of the game. The chosen value would subsequently be applied whenever she did not submit an active decision in a given round. This is the 'self-nudge'. The design of the self-nudge ensures, in contrast to alternative designs, a faithful implementation of Thaler and Sunstein's thought experiment of asking subjects to deliberate on their preferred default. It also ensures clean comparability across treatments: In all three default conditions, nudges are unchangeable across rounds and are not influenced by prior experience.[8]

We deliberately assembled, through random sampling, a higher number of experimental groups in SELF to optimize power given the expected higher variance in the endogenously chosen default values compared to the SELFISH and SOCIAL. In particular, we aimed at having 20 independent groups in each of the exogenous SELFISH and SOCIAL treatments, and 30 independent groups in the SELF treatments.

Voluntary contributions can be embedded in different information structures (see (e.g. Neugebauer et al., 2009; Angelovski et al., 2018)). We focus on four main types, depending on whether defaults are kept private (PvD) or their average disclosed to all group members (PuD) and on whether contributions are kept private (PvC) or their average (and active support) made public to all group members (PuC). In both cases, 'private' means that group members only know they own default (PvD) or contribution (PvC). 'Public' means that group members learn other group members' average default (PuD) or average contribution after every round, plus the number of active contribution decisions (PuC) among other group members.

[8]Alternative designs for a self-nudge could involve providing experience with the VCM mechanism (under some other default), allowing one or more opportunities to change the default, allowing an opt out of being able to change the default, different information structures and many other features. By sacrificing comparability with exogenous nudges, these alternatives are a natural next step in future research.

Table 1: Treatments and number of observations

	Number of subjects (groups)			
Feedback information:	PvD-PuC	PvD-PvC	PuD-PvC	PuD-PuC
Perfect free-riding (SELFISH)	72(18)	68(17)	76(19)	80 (20)
Full cooperation (SOCIAL)	80 (20)	80 (20)	84(21)	80 (20)
Self-determined (SELF)	112 (28)	116 (29)	116 (29)	116 (29)
Total	264 (66)	264 (66)	276 (69)	276 (69)

Note: The entries in the table denote the number of subjects (groups) per condition. A total of 1,080 subjects (in 270 independent groups) participated in the experiment. "PvD-PuC" denotes a treatment with private defaults and public contributions (the standard information structure), "PuD-PvC" one with public defaults and private contributions, and so on.

2.2 Online recruitment

Subjects were recruited using an Internet polling company. Panel members who agreed to participate in an "interactive survey" lasting ten rounds over ten days entered basic demographic information during recruitment. In total, 1,651 panel members pre-registered for participation and were randomly assigned to one of the twelve treatments. 1,156 panelists responded to our invitation email for round one by signing into the experimental website and completed round one. Since responses did not allow to form complete groups of four in every treatment, seven subjects were dismissed in exchange for a fixed compensation of \$5 as pre-announced. This left a sample of 1,080 subjects.[9] There is little evidence for systematic selection of pre-registered panelists into the experiment based on the sociodemographic characteristics available to us from recruitment. An exception is the show-up rate among female registrants, which is about six percentage points lower than males (see Table A.1 in the Appendix). Tables A.2 and A.3 in the Appendix reports summary statistics for the full sample and by treatment and shows that observable characteristics are balanced across treatments overall.[10] For additional robustness, our parametric results control for

[9] A common concern for online experiments is a loss of control about subjects identity and hence, multiple participation. The polling company prevented double registration with the same panel IDs, which is confirmed by our data. What remains is the possibility of subjects using multiple accounts with the polling company. We have data on subjects' IP addresses for rounds 9 and 10 of the experiment. Of the 1,080 subjects, there were five (six) pairs of observations with identical IP addresses in round 9 (10). Of those, four pairs had identical IP addresses in both rounds. Among those, one pair of observations had stated the same region of residence, ZIP code, gender, education, and income level upon registration. This leaves up to 12 out of 1,088 observations that exhibit some evidence for potential identity duplicates, among which two observations that are extremely likely to represent the same subject.

[10] Joint significance F-tests yield that gender, education, and income do not differ across treatments. For age bracket, the treatments differ statistically (p=0.029), but not substantively. The largest difference across treatments is not more than 0.7 points of a ten-year age

demographics.

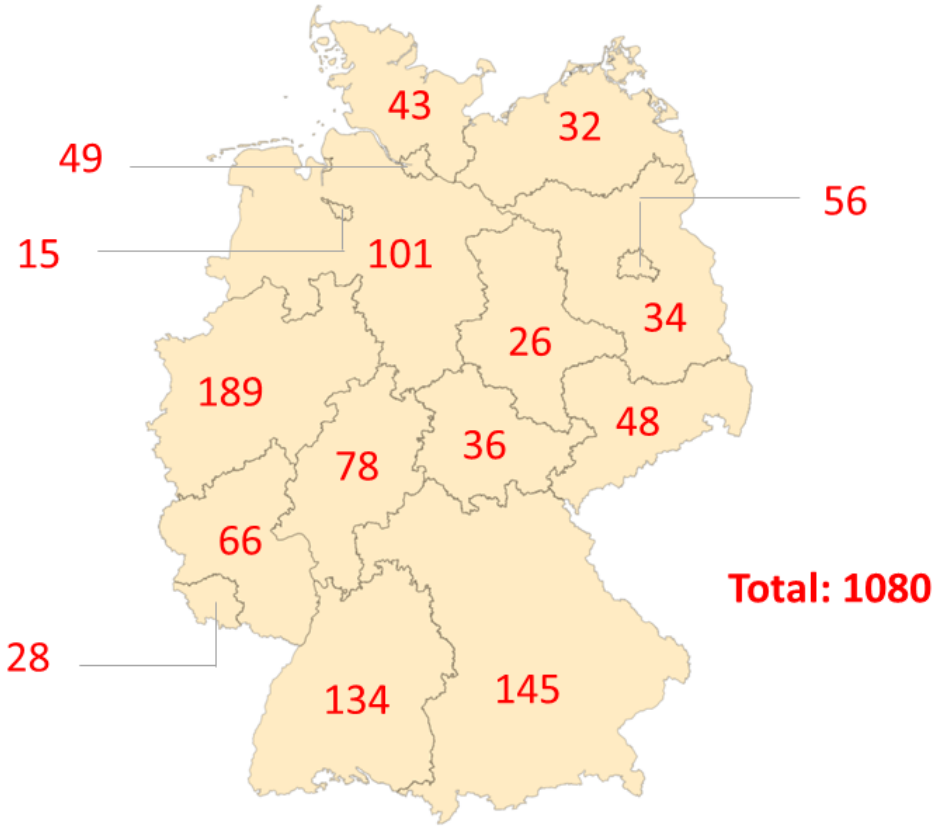


Figure 1: Distribution of participants across the country by state.

2.3 Experimental procedure

Each experimental round commenced with an invitation email sent out early in the morning that contained login information and the link to the experimental website (Figure 2). On the login screen of the experimental website, subjects had to manually enter the login information, that is, a user name and password. Intentionally, login credentials could not be saved in the browser in order to maintain some effort cost of participation.

In the first round only, after logging in, subjects received an introduction with a short explanation of the VCM, including information about the within-subjects random incentive scheme (one round was randomly drawn for payment) and the conversion rate of the experimental currency.[11]

bracket.

[11]The instructions could be reviewed later on any screen and in any round by clicking on a link available in the northeastern corner of the screen.

The following main decision screen, presented in every round, displayed a history of play, provided access to a "payoff calculator" tool, reminded subjects of their own default contribution of their treatment (from round two), and elicited contribution decisions. The history of play showed, for each previous round, own previous contribution decisions. When the treatments included public defaults (PuD), the decision screen also reminded subjects of the average default contribution setting of the other three group members. When the treatment included public contributions (PuC), the decision screen also provided the history of average contributions of the other three group members to the group account and the number of actively submitted decisions in the group. The "payoff calculator" allowed subjects to learn about the payoff consequences of different allocations. Subjects made their decisions how to allocate their per-round endowment between their private account and the group account using two fields, one for each account, that featured an auto-completion function to ensure that all of the endowment was used.

Depending on the experimental round and treatment, there were between two and four more screens. In the first round only, after making their contribution decisions, subjects were informed about their specific non-participation default rule. For the two exogenous default treatments, the screen simply explained the procedure. Subjects assigned to the SELF treatment were to choose their own default contribution. In the course of the experiment, subjects were reminded of their default contribution, either exogenous or endogenously chosen, in each invitation email and on the decision screen of each following round. In the SELF treatments, the next screen elicited subjects' beliefs about the average default contribution of the other members in the group. In all treatments and rounds, the round concluded with a screen where subjects had to state their beliefs regarding their other group members' participation and contributions in that round. The belief elicitation was not incentivized (Gächter and Renner, 2010; Charness et al., 2021).

Each round was closed at about 2:00 AM the following day. After the final round, the experimenters randomly drew the payoff-relevant round, payoffs were computed, and payments initiated through the online polling company's payment infrastructure. The currency we used in the experiment was the same currency that the online polling company used to incentivize their surveys (1 unit = \$0.05). On average, subjects earned about 101 units of the experimental currency (i.e., \$5.05).

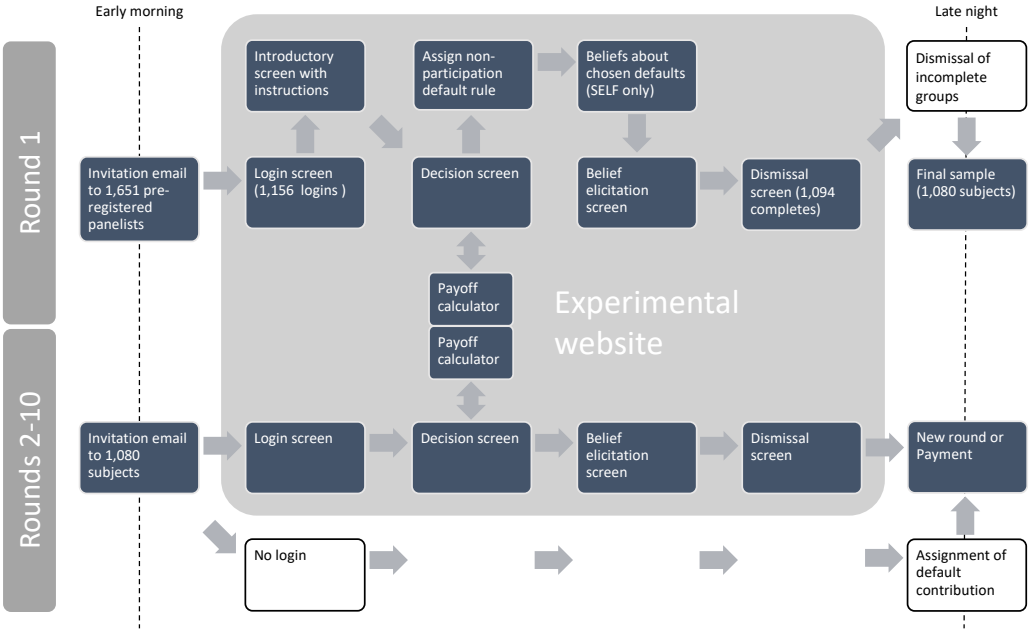


Figure 2: Flowchart of experimental procedure

3 Results – Standard Case

Private defaults and public contributions characterize what could be regarded as the standard case for the information structure under which people contribute to public goods. This section therefore starts with comparing default rules for the standard case (PvD-PuC). Section 4 examines whether these results generalize to the other three information structures and Section 5 presents exploratory analysis on defaults choices and beliefs that could enhance our understanding of the underlying mechanisms and motivate future research.

Figure 3 presents the distribution of self-determined default contribution levels set by subjects in the SELF default rule. The modal and median choice under the SELF default rule is exactly half of the endowment (40 experimental currency units), chosen by 33.9 percent of subjects. By contrast, only about 7.1 percent of the subjects chose either a default of zero contribution or a default of full contribution. We can thus formulate our first result.

Result 1 *When asked to set their own non-participation default contribution, subjects' modal and median choice was to equally split the endowment. On average, subjects set a default contribution of 44 percent of their endowment. Only about 7 percent of subjects each set the default contribution at zero or 100 percent of their endowment.*

To our understanding, this is the first reported evidence on how individuals

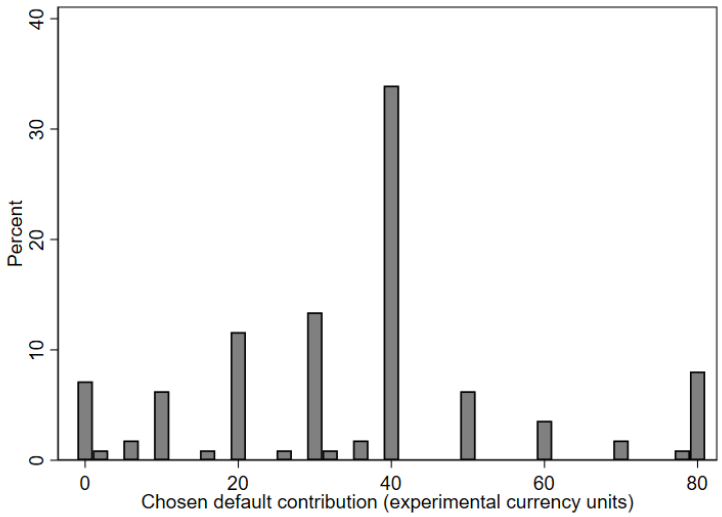


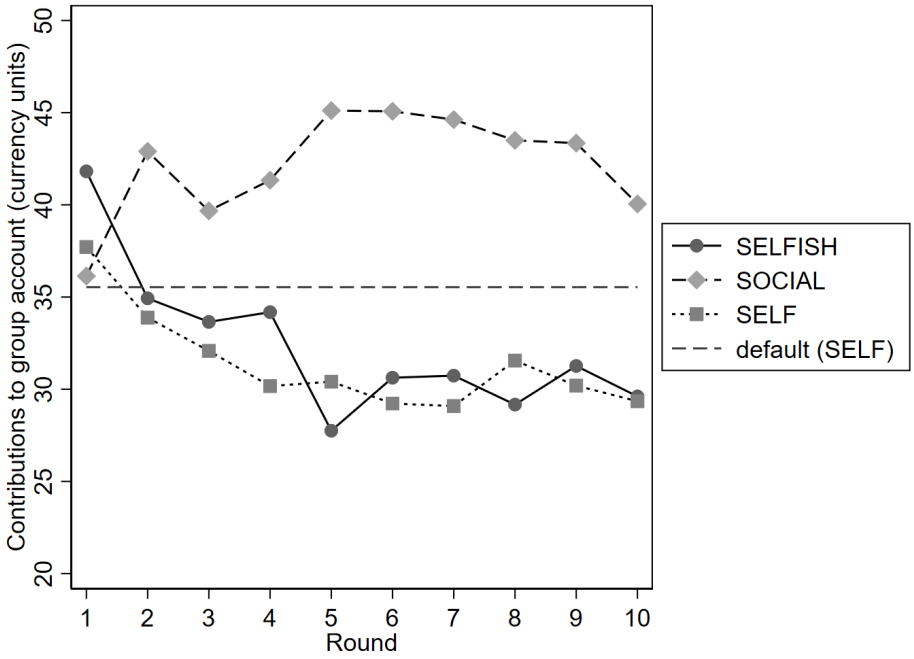
Figure 3: Histogram of chosen non-participation default contribution values in the SELF default rule (PvD-PuC information condition).

set their own nudge in a social dilemma setting. It shows that for 93 percent of subjects, either the most common default in VCMs, namely contributing zero, or the default in line with maximizing group benefits, namely contributing the full endowment, differs from what they would choose themselves. Concerns about a possible misalignment between a choice architect and the targeted individuals therefore have an empirical basis.

Moving on to contributions in the VCM, we first focus on the two exogenous default treatments. Figure 4 shows average total contributions over the course of the experiment for each of the three default conditions. For the baseline of a zero contribution default (SELFISH), total contributions start out at 41.8 units (52.3 percent) in round 1. We observe a decline over time, ending up at 29.6 units (37 percent). Hence, behavior in SELFISH is similar to the typical pattern of play in standard VCM experiments (e.g., Dawes and Thaler, 1988; Fehr and Gächter, 2000). By contrast, total contributions in SOCIAL, while starting at about the same value as SELFISH before the introduction of defaults (36.1 units or 45 percent), first increase and then remain stable over time, ending at 40.1 unit (50 percent of the endowment). Compared to SELFISH, contributions in the SOCIAL treatment are therefore around a third higher, relatively speaking.

Result 2 *Under a social nudge (full contribution default), total contributions are around 30 percent higher than under a selfish nudge (zero contribution default).*[12]

[12]Increase computed on the basis of relative differences in contributions between treatments,



Note: The horizontal dashed line denotes the average self-determined default contribution in SELF.

Figure 4: Evolution of average total contributions (PvD-PuC information condition)

A formal analysis based on random-effects GLS regressions reiterates Result 2 . Table 2 reports the treatment effect of SOCIAL in units contributed, with SELFISH as the baseline default rule.[13] An alternative to is to estimate linear models with a two-way robust standard errors clustered across (i) groups and (ii) rounds. For robustness, in Appendix B we conducted such regressions for each panel data regressions presented in the paper. The results of the panel-data random-effect regressions and the pooled regressions with a two-way robust standard errors are similar. Model 1 in Table 2 shows the random-effects regression results for total contributions by default condition, a continuous round variable, and the interaction of both, controlling for several sociodemographic characteristics and the weekday of the round. These results estimate total contributions under the social nudge to be 9.6 units (12 percentage points) higher than under the selfish nudge ($p = 0.009$). This corresponds to a 26 percent increase under SOCIAL relative to SELFISH. We observe no significant time

averaged over 10 rounds. Raw data available in Table A.4.

[13] A Breusch-Pagan LM test for random effects rejects the Null that there are no significant differences across subjects, hence the RE GLS is preferred over a simple pooled OLS for the panel data. For an equivalent pooled OLS with a two-way clustered robust standard errors see Appendix B.

trend under the social nudge (Wald test, $p = 0.1737$), but the typical negative trend under the selfish nudge ($p = 0.023$): The difference in cooperation between the exogenous default rules is diverging over time.

Result 2 can be explained by at least two candidate mechanisms. At the extensive margin of contributing, default contributions by non-participating subjects "mechanically" drive total contributions apart because every non-participating subject under SELFISH contributes zero, while every non-participating subject under SOCIAL contributes the entire endowment. Non-participation is infrequent, but not negligible: The average share of non-participating subjects is 11.1 percent in the SELFISH treatment and 14.7 percent in SOCIAL. This implies that the 'mechanical effect' at the extensive margin is responsible for a significant share of the difference in total contributions.[14] At the same time, the mechanical effect cannot explain differences in dynamic play. Beyond the mechanical effect, behavioral feedback at the extensive margin could further affect total contributions. The experimental evidence, however, fails to support the conjecture that treatments cause differences in subjects' propensity to actively participate. In particular, the regression results shown in Model 3 underline that participation does not differ significantly, either in levels or in the trend, between the two exogenous treatments.

The second candidate mechanism is behavioral feedback at the intensive margin of contributions. Model 2 shows that, for levels, the data are not supportive of this conjecture. On average, contribution levels by active contributors look very similar across the two exogenous default treatments. Treatment averages in SELFISH and SOCIAL are similar and not considerably different (t-test, $p = 0.582$). Behavioral feedback at the intensive margin could, however, have dynamic effects: At the mean experimental round, we estimate a significant trend effect. Contributions in SELFISH decrease by an estimated 0.96 units per round (t-test, $p = 0.019$) but they do not decrease in SOCIAL (Wald test, $p = 0.611$), explaining the differences in the trend observed in total contributions. The comparison of SELFISH and SOCIAL shows that variations in exogenous defaults can cause variations in contribution levels and trends through mechanical and behavioral effects at the extensive and intensive margin.

In the SOCIAL and SELFISH treatments, subjects make participation and contributions choices under exogenously imposed defaults. In the SELF treatment, they do so under self-determined nudges. Revisiting Figure 4, the pattern of total contributions under the self-nudge is very similar to that under the

[14] For reference, a five percentage point increase in participation in the SOCIAL as opposed to the SELFISH treatment automatically raises total contributions by 4 units.

Table 2: Determinants of average contributions and participation, standard case

	Total contributions (1)	Active contributions (2)	No participation (3)
SOCIAL	9.595***	1.977	0.027
	(3.674)	(3.588)	(0.033)
SELF	-0.614	-3.151	0.020
	(3.393)	(3.427)	(0.036)
Round (mean centered)	-0.988**	-0.955**	-0.004
	(0.436)	(0.407)	(0.004)
SOCIAL × Round	1.428***	1.093**	0.002
	(0.538)	(0.488)	(0.006)
SELF × Round	0.299	0.228	0.006
	(0.487)	(0.465)	(0.005)
Constant	36.544***	39.194***	0.132*
	(5.324)	(4.763)	(0.069)
Additional Control	Yes	Yes	Yes
Observations	2470	2209	2223
R-squared (overall)	0.073	0.055	0.053

Note: Random-effects GLS regressions. Dependent variable is average contributions in units of endowment. Robust standard errors are clustered for experimental groups. Baseline is the SELFISH default condition. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. Additional controls include gender, age, region of residence, education, income, weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. Model 3 is based on a linear probability model. *, **, and *** denote significance level of at least 10%, 5%, and 1%, using two-sided tests throughout.

selfish nudge and considerably lower than under the social nudge. The random-effects GLS regression in Model 1 of Table 2 confirms this: Contributions do not significantly differ between SELF and SELFISH, both in level and in trend. On the other hand, total contributions are significantly higher in SOCIAL than in SELF (Wald tests, in levels, $p = 0.001$; and trend, $p = 0.005$). This gives rise to result 3.

Result 3 *Total contributions under self-nudges are not significantly higher or lower than under a selfish nudge of perfect free-riding. But compared to total contributions under a social nudge of full cooperation, they are about 26 percent lower.*[15]

An examination of the underlying effects shows that at the extensive margin, non-participation in SELF does not differ statistically from that in the SELFISH and SOCIAL treatments (see Model 3 of Table 2). The mechanism of default contributions implies that similar shares of non-participants across treatments should place total contributions in SELF between those in the exogenous treatments. Average default contribution in SELF are, after all, about 36 units (44 percent) compared to 0 units in SELFISH and 80 units in SOCIAL. Yet, this is not the case. The reason is that an intensive margin effect of self-nudging exerts a downward pressure on active contributions: Active participators in SELF contribute 5.16 units less than those in SOCIAL, a marginally significant difference (Wald test, $p = 0.088$). Active contributors in SELF also contribute, on average, 3.15 units less than those in SELFISH, but the difference is not significant (t-test, $p = 0.358$). Yet, the higher average intensive margin contributions under SELFISH offsets the "mechanical effect" of defaults in SELF, leading to no significant difference in total contributions between these default rules.

4 Generalizability

While private defaults and public contributions suggest themselves as the most common information structure, other information structures could be present by nature of the social dilemma or by construction. This raises the question whether results 1 through 3 generalize to other cases of information structure.

Starting, as before, with the question of ethicality, we find that the modal and median choice in all treatments is exactly 40 experimental currency units

[15]Decrease computed on the basis of relative differences in contributions between treatments, averaged over 10 rounds. Raw data available in Table A.4.

(half of the endowment). This was chosen by 34 (PvD-PuC), 31 (PvD-PvC), 31 (PuD-PvC), and 33 (PuD-PuC) percent of subjects. In addition, only 7.1 (7.1), 6.0 (6.0), 3.4 (5.2), and 4.3 (7.8) percent of the subjects chose a default of zero (full) contribution in the respective PvD-PuC, PvD-PvC, PuD-PvC, and PuD-PuC treatments.[16] The self-determined default choices do not differ significantly across the four structures (Kruskal–Wallis H test, $p = 0.6229$). This leads to the conclusion that irrespective of the information about defaults and contributions, subjects' average self-determined default choice is close to half of the endowment. This provides additional support to the earlier claim that the exogenous defaults of the typical VCM experiment and those of most public goods games coincide with default preferences for no more than a small part of the population.

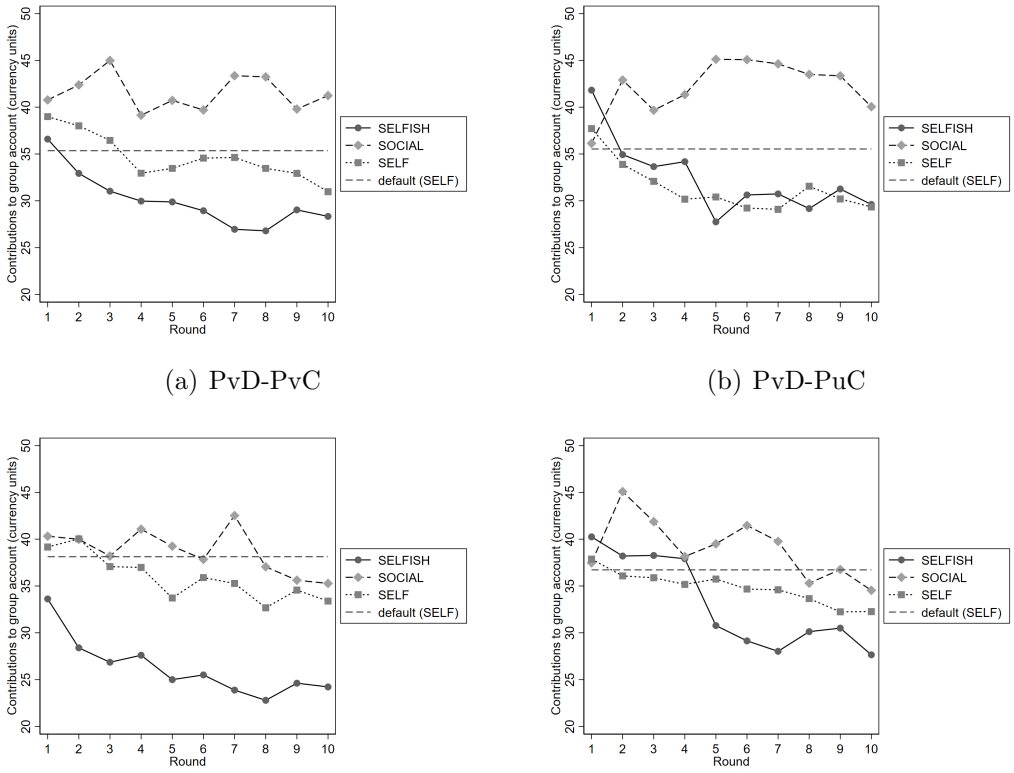
Result 4 *Result 1 generalizes to all four information structures: When asked to set their own non-participation default contribution, subjects' modal and median self-nudge was an equal split of the endowment. Only between 3 and 8 percent of subjects each set the default contribution at zero or 100 percent of their endowment.*

Moving on to efficiency, Figure 5 shows the evolution of average total contributions in the three defaults rules for each information condition. This figure visually foretells the three econometric conclusions below: First, a social nudge yields considerably higher average total contributions than a selfish nudge, irrespective of the information condition. Second, when defaults are private, the social nudge yields considerably higher total contributions than a self-nudge. Public defaults attenuate the difference in average total contribution between a social and a self-nudge, however.

Table 3 shows panel regression models on the effect of the default conditions on the total contributions, active contributions, and participation, controlling for the feedback information, the round, and additional controls (e.g., gender, age, region of residence, education, income, weekday of the experiment). In estimating the models, we set SELFISH as the baseline default rule and PvD-PuC as the baseline information condition. Model 1 shows that total contributions are significantly higher in SOCIAL than in SELFISH in all information conditions (at $p = 0.012$ (PvD-PuC), $p = 0.002$ (PvD-PvC), $p = 0.0005$ (PuD-PvC), and $p = 0.070$(PuD-PuC) - all reported tests are two-sided).[17] This generalizes

[16]Figure C.1 in the Appendix presents the distribution of default contribution levels self-determined by subjects (in SELF) under all four information structures.

[17]In particular, the p-value in PvD-PuC information condition is based on a t-test, the values in the remaining information conditions on a joint-significance Wald test.



(a) PvD-PvC

(b) PvD-PuC

(c) PuD-PvC

(d) PuD-PuC

Note: Horizontal dashed lines denote the average self-determined default contribution in SELF.

Figure 5: Evolution of average total contributions

the results from Section 3, albeit with less effect strength when defaults and contributions are public).

Result 5 *Result 2 generalizes to all information structures: Total contributions are significantly higher under a social nudge of a full cooperation default than under a selfish nudge of a zero contribution default.*

Turning to the self-nudges, we next compare total contributions in the SELF and SELFISH treatments. When contributions are public (PvD-PuC and PuD-PuC), total contributions are not significantly higher in SELF than in SELFISH (t-test, $p = 0.649$; Wald test, $p = 0.537$). However, when contributions are private and defaults are public, total contributions under SELF are significantly higher than under SELFISH (Wald test, $p = 0.003$).

Comparing total contributions in the SELF and SOCIAL treatments, the social nudge leads to higher levels than the self-nudge when defaults are private (joint-significance Wald test, $p = 0.0006$ (PvD-PuC), $p = 0.012$ (PvD-PvC)). However, when defaults are public, total contributions are not signif-

Table 3: Determinants of contributions and participation

	Total contributions (1)	Active contributions (2)	No participation (3)
SOCIAL	7.979**	0.642	0.041
	(3.181)	(3.121)	(0.029)
SELF	-1.279	-4.111	0.035
	(2.808)	(2.869)	(0.028)
Public Defaults (PuD)	-1.329	-1.136	0.020
	(2.687)	(2.737)	(0.024)
Private Contributions (PvC)	-4.058	-3.762	0.024
	(2.654)	(2.696)	(0.024)
SOCIAL × PuD	-1.849	-1.060	-0.041
	(3.695)	(3.617)	(0.032)
SOCIAL × PvC	4.029	4.218	-0.028
	(3.647)	(3.562)	(0.032)
SELF × PuD	3.040	2.819	-0.049
	(3.300)	(3.365)	(0.031)
SELF × PvC	6.375*	6.072*	-0.017
	(3.296)	(3.352)	(0.031)
Round (mean centered)	-1.040***	-0.793***	0.004*
	(0.156)	(0.148)	(0.002)
SOCIAL × Round	0.838***	0.378*	-0.006**
	(0.237)	(0.214)	(0.003)
SELF × Round	0.360**	0.058	-0.001
	(0.183)	(0.181)	(0.003)
Constant	34.775***	38.668***	0.090***
	(3.204)	(3.215)	(0.031)
Additional controls	Yes	Yes	Yes
Observations	10120	9210	9108
R-squared (overall)	0.052	0.029	0.038

Note: Random-effects GLS regression. Robust standard errors are clustered for experimental groups. Baseline is the SELFISH default rule. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non-participation. "SELF" denotes a treatment with self-determined default contributions. "Public Defaults (PuD)" denotes a treatment in which average default contributions are revealed to the group. "Private Contributions (PvC)" denotes a treatment in which contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, weekday of the experiment. The Round variable is mean-centered: Coefficient estimates correspond to marginal effects estimated at mean experimental round. Model 3 is based on a linear probability model. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

icantly higher in SOCIAL than in SELF (Wald test, $p = 0.416$ (PuD-PvC), $p = 0.151$ (PuD-PuC)). Finally, total contributions decline over time in SELFISH and SELF ($p < 0.0001$ for both), but no significant time trend is observed in SOCIAL ($p = 0.264$).

Result 6 *Result 3 does not generalize to all information structures. When defaults are public, total contributions under a social nudge are not significantly*

higher than under self-nudging. When defaults are public and contributions are private, total contributions under self-nudging are significantly higher than under a selfish nudge and not significantly lower than under the social nudge.

Model 2 informs about the determinants of active contributions, that is decisions by members participating on the day. The estimates in Table 3 indicate that overall there are no difference in active contributions between the treatments.[18] In addition, there is a decline in active contributions over time in all treatments ($p < 0.0001$ in both SELFISH and SELF, $p = 0.008$ in SOCIAL).

Finally, Model 3 shows the determinants, by way of a linear probability model, of participation, that is the share of members actively contributing on the day. There are no significant differences in participation between the default rules across the information structure and no evidence that information structures affect participation.[19]

Result 3 should not be interpreted to mean that the trade-off between efficiency and ethicality is resolved when defaults are made public. Comparing across information structures, for example, shows that the group-level efficiency induced by the social nudge under private defaults and public contributions was higher than the efficiency induced by the self-nudge under public defaults and private contributions (Wald test, $p = 0.053$). A group that would have undertaking measures in our experiment to deviate from the standard information structure in order to make defaults public would therefore not have reached a higher efficiency overall, even if the measures were costless.

5 Defaults Setting and Beliefs

The purpose of the experimental design was to examine how subjects set their own nudges in a paradigmatic social dilemma setting when given the opportunity, and whether cooperation in social dilemmas differs as a result of the different choice architectures of exogenous and endogenous nudges. With the results in hand, we report here on additional exploratory evidence in order to provide guidance for future research on the underlying causes. We first report on the behavioral correlates of endogenous default choices before reporting on

[18] Two exceptions with marginal significance are that active contributions in PvD-PuC are marginally higher in SOCIAL than in SELF (Wald test, $p = 0.069$), and that active contributions in PuD-PvC are marginally higher in SELF than in SELFISH (Wald test, $p = 0.089$).

[19] Figures C.2 and C.3 in the Appendix show the intensive marginal contributions and non-participation rates.

the evolution of beliefs in the different treatments.

In SELF, subjects choose their own default contribution. This choice is very strongly correlated with the contribution decision in round 1.[20] A majority of the subjects, between 62 and 72 percent depending on the information structure, chose exactly the first period contribution as default. This would reaffirm our conjecture that self-nudges represent subjects' preferences: Subjects set defaults close to how they decide when they participate. Among the minority whose self-nudge deviated from their first-round contributions, the majority (56 to 74 percent) chose a value that was smaller than their first-round choice. This could reflect subjects' beliefs about the future evolution of group contributions, but other cognitive and affective drivers could play a role (Bouwmeester et al., 2017; Goeschl and Lohse, 2018).

Information structures may affect subjects' beliefs and expectations (Neugebauer et al., 2009; Angelovski et al., 2018). For example, learning about other group members' defaults could affect subjects' beliefs about others' likely contribution, but also provide information about a descriptive norm in the group. These beliefs and expectations in turn affect behavior of conditional cooperators in the group (e.g., Keser and Van Winden, 2000; Fischbacher et al., 2001; Fischbacher and Gächter, 2010). To investigate this channel, the experiment elicited (non-incentivized) subjects' beliefs about others' behavior. Table 4 shows a series of random-effect panel data regressions on the beliefs in the different treatments. The elicitation item for Model 1 was the belief on "how many other group members did not participate in this round." The results indicate that there is no difference between treatments regarding beliefs about the non-participation rate. Subjects do not appear to expect systematic connections between default rules and the participation rate,[21] but expect a positive link between contributions being public and active participation (t-test, $p = 0.010$). In addition, subjects do not expect participation to evolve over time. The elicitation item for Model 2 was the belief on "how many other group members participated *and* gave more than zero." Results indicate that public contributions increase subjects' expectations that more group members actively participate and contribute a positive amounts (t-test, $p = 0.000$). When defaults are public, subjects in the self-nudge treatment expect to encounter more active participants who are contributing than in SELFISH or SOCIAL (Wald test,

[20] Correlation coefficients are 0.673 (PvD-PuC), 0.794 (PvD-PvC), 0.925 (PuD-PvC), and 0.943 (PuD-PuC).

[21] The only difference is that we find that in the PuD-PuC information condition, subjects expect less non-shows in SELF than in SELFISH and SOCIAL (Wald tests, $p = 0.0138$ and $p = 0.0135$, respectively.)

SELF vs. SELFISH, $p = 0.0002$ (PuD-PvC), $p = 0.0913$ (PuD-PuC); SELF vs. SOCIAL, $p = 0.0008$ (PuD-PvC), $p = 0.0010$ (PuD-PuC)). Finally, subjects expect a decline in the number of group members who participate and make a positive contribution ($p < 0.0001$ in SELFISH and SOCIAL, $p = 0.010$ in SELF).

The elicitation item for Model 3 was the belief on "the average contribution by others who actively participated and gave more than zero". Table 4 shows that public contributions increase subjects' expectation about non-zero contributions of active participators (t-test, $p = 0.003$). In addition, specifically when defaults are public and contributions private (PuD-PvC), subjects expect higher average active intensive margin contributions when subjects self-nudge rather than facing a selfish or social nudge ($p = 0.0004$ and $p = 0.010$, respectively; no significant difference between SELFISH and SOCIAL). This evidence on beliefs aligns with the evidence on contributions: It is in the PuD-PvC information structure that self-nudges outperformed the selfish nudge and were no less efficient than the social nudge. We also find that in SELFISH, subjects expect a decline over time of intensive-margin contributions from those active participators (t-test, $p = 0.006$). Such a negative trend is not observed in SOCIAL and SELF (Wald tests, $p = 0.394$ and $p = 0.933$, respectively). Finally, for Model 4, we constructed a variable that combines the latter two belief elicitation items to measure a "belief on average contribution by others who actively participated". This model returns very similar results as Model 3. Public contributions increase the expected amount to be contributed by active participants (t-test, $p = 0.001$). In the specific case of PuD-PvC, expected active contributions are higher in SELF than in SELFISH and SOCIAL ($p = 0.0015$ and 0.0033, respectively; no significant difference between SELFISH and SOCIAL). Finally, in SELFISH we observe a decline over time in expected contributions from those active participators (t-test, $p = 0.025$). However, such a decline in expectations over time is not observed in SOCIAL and in SELF (Wald tests, $p = 0.181$ and $p = 0.590$, respectively).

Our motivation to elicit beliefs is that we expected them to be the channel relating the different information conditions to the actual behavior. Explicitly, we hypothesized that the different feedback information conditions lead to different expectations about cooperation, resulting in different active and total contributions. If this is true, then when we include a 'beliefs' variable in the estimations of main results (Table 3), we expect to find this variable to significantly affect contributions, but at the same time to "turn off" the partial effect

Table 4: Determinants of beliefs

	Beliefs about nonpart. rate (1)	Beliefs about rate pos.ctr. (2)	Beliefs about pos. contrib. (3)	Beliefs about active contrib. (4)
SOCIAL	-0.010	-0.015	3.271	3.098
	(0.031)	(0.034)	(2.443)	(2.453)
SELF	-0.044	-0.008	-1.032	-1.272
	(0.028)	(0.029)	(2.091)	(2.147)
Public Defaults (PuD)	-0.030	0.000	2.672	2.563
	(0.024)	(0.024)	(1.865)	(1.858)
Private Contributions (PvC)	0.061***	-0.135***	-5.578***	-6.331***
	(0.023)	(0.024)	(1.846)	(1.828)
SOCIAL × PuD	0.010	-0.032	-5.841**	-6.334**
	(0.033)	(0.035)	(2.734)	(2.754)
SOCIAL × PvC	0.007	0.062*	3.450	3.799
	(0.033)	(0.035)	(2.739)	(2.761)
SELF × PuD	-0.013	0.052*	1.204	1.738
	(0.029)	(0.031)	(2.496)	(2.535)
SELF × PvC	0.033	0.059*	6.139**	6.531**
	(0.028)	(0.031)	(2.514)	(2.549)
Round (mean centered)	-0.002	-0.007***	-0.334***	-0.288**
	(0.002)	(0.002)	(0.122)	(0.128)
SOCIAL × Round	0.002	-0.002	0.453**	0.483**
	(0.003)	(0.003)	(0.184)	(0.193)
SELF × Round	0.002	0.003	0.326**	0.342**
	(0.003)	(0.003)	(0.155)	(0.162)
Constant	0.228***	0.816***	31.043***	29.508***
	(0.033)	(0.034)	(2.200)	(2.258)
Additional controls	Yes	Yes	Yes	Yes
Observations	8135	8876	8647	8647
R-squared (overall)	0.051	0.067	0.049	0.052

Random-effects GLS regressions. Robust standard errors are clustered for experimental groups. Baseline is the SELFISH default rule. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. "Public Defaults (PuD)" means that average default contributions are revealed to the group. "Private Contributions (PvC)" means contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, and the weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. All reported tests two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

of the feedback information (as feedback information affects indirectly through subjects' beliefs and expectations). Table 5 shows the estimation results of random effect panel-data models where we now include beliefs about active contributions. In particular, we use the constructed belief variable measuring "belief on average contribution by others who actively participated" (see Model 4 of Table 4). Since we only elicited belief from those subjects who actively participated in each round, we can only estimate the effects on (1) total con-

tributions (i.e, contributions including those made by non-participants through their default choice), and (2) active-participation contributions.

Table 5: Total contributions and active contributions

	Total contributions (1)	Active contributions (2)
Beliefs about active contrib.	0.463***	0.464***
	(0.024)	(0.024)
SOCIAL	-0.291	-0.315
	(2.276)	(2.278)
SELF	-2.871	-2.888
	(2.126)	(2.128)
Public Defaults (PuD)	-1.576	-1.526
	(2.102)	(2.111)
Private Contributions (PvC)	0.008	-0.041
	(2.089)	(2.097)
SOCIAL × PuD	1.061	1.012
	(2.714)	(2.720)
SOCIAL × PvC	1.679	1.736
	(2.654)	(2.662)
SELF × PuD	1.481	1.428
	(2.480)	(2.488)
SELF × PvC	2.304	2.355
	(2.497)	(2.505)
Round (mean centered)	-0.711***	-0.724***
	(0.121)	(0.120)
SOCIAL × Round	0.231	0.243
	(0.177)	(0.177)
SELF × Round	0.011	0.023
	(0.146)	(0.145)
Constant	25.241***	25.295***
	(2.499)	(2.504)
Additional controls	Yes	Yes
Observations	8647	8646
R-squared (overall)	0.314	0.314

Random-effects GLS regressions. Robust standard errors are clustered for experimental groups. Baseline is the SELFISH default rule. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. "Beliefs about active contrib" is constructed from belief elicitation items 2 and 3, indicating the average contribution of active participators. "Public Defaults (PuD)" means that information about average defaults is revealed to the group. "Private Contributions (PvC) " means that contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, and the weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

Table 5 indicates that stated beliefs significantly explain total contributions (Model 1, $p = 0.000$) and active contributions (Model 2, $p = 0.000$), while shutting down the effects of the feedback information conditions on the default

rules. Also, when controlling for the beliefs, total and active contributions decline over time in all default rules ($p \leq 0.0002$ for all default rules).

Thus, taken together, the results in Tables 4 and 5 can explain cooperation behavior in the different treatments. By and large subjects do not expect that the three default rules affect non-show (see Model 1 in Table 4). Indeed non-shows are not different between treatments (see Model 3 in Table 3). This leads to a "mechanical effect" of highest total contributions in SOCIAL, middle in SELF, and lowest in SELFISH. Second, Model 3 and 4 in Table 4 indicate that subjects expect lower contributions in SOCIAL when defaults information is public and higher contributions in SELF when contributions information is private. This can explain why – when defaults are public and contributions private – conditional cooperators actively contribute more in SELF than in SOCIAL, which offsets the "mechanical effect" from the defaults in favor of the social nudge. In the standard information structure of PvD-PuC, SOCIAL results in higher cooperation than SELF due to the mechanical effect. Finally, Table 5 indicates that the beliefs may be the channel through which feedback information affects contributions.

6 Conclusions

Are self-nudges an answer to the criticism that has been leveled at the practice of deploying nudges in the context of social dilemmas? This question merit investigation as social nudges become increasingly popular, spreading beyond contexts for which the nudge concept was originally intended and thereby attracting objections on ethical grounds. Self-nudging, i.e. giving individuals control over how they will be nudged for social ends, is one candidate answer to resolving a possible conflict between choice architect and targeted individual. Our study investigated experimentally what happens when individuals are put into the role of choice architect for their own future decisions in a social dilemma setting. For the social dilemma, we used the voluntary contribution mechanism, an environment that is conducive to quantitative measurement of default setting, behavior, and outcomes. Our experiment studied two main outcomes under one of three choice architectures, self-nudges chosen by the participants, the common selfish nudge of contributing nothing, or a social nudge of contributing the entire per-round endowmwnet, and under one of four information structures, with defaults and contributions private or public information. The first main outcome is how the self-nudges are set. This allows to understand more about

the extent to which externally set defaults violate individuals' own preferences. The second main outcome is the efficiency of the group in the social dilemma. To give the idea of nudging meaning and purpose, the experimental procedures relied on a multi-session online environment in which subjects from the general population have to overcome a minimum of inertia to take a daily contribution decision repeatedly.

We find, first, that the conflict between the choice architect and the targeted individual is real: When subjects chose their default contributions themselves, heterogeneity in self-nudges was high, ranging from zero to full contribution. The average default was just under half of the endowment, with an equal split the modal and median choice. Only a small minority of subjects chose either the selfish or social nudge for themselves. Our evidence shows, in other words, that when she was actually "asked", Thaler and Sunstein's "reflective individual" set her default at a very different level from the standard benchmarks used in social dilemma experiments. This discrepancy provides empirical support to arguments that social nudges violate the philosophical premises and libertarian ethics of nudging. Libertarian paternalists favoring the social nudge would need to justify why that nudge is allowed to override the nudges that people would choose for themselves. Incidentally, the fact that our subjects rarely opted for the zero contribution default provides empirical support for arguments that most public goods experiments in research, including our own, have the 'wrong' choice architecture.

We also find that the social nudge, while ethically problematic, delivers efficiency in social dilemmas: Groups in which full contribution was the non-participation default contributed more in the VCM than groups with zero contribution defaults. The reason for why the social nudge worked in our experiment can be satisfactorily explained by the mechanical, rather than behavioral effects. This lack of a detectable link between social nudging and behavior is worth stressing. For one, it would imply that finding evidence for "psychological reactance" is likely to require other designs. The absence of a link also contrasts with previous evidence that exogenous defaults can affect or alter the social norm. This link may therefore merit another look in future research.

The third, and perhaps most important, insight is that the trade-off between ethicality and efficiency of nudging can be mitigated, but not be reconciled. In the standard case of private defaults and public contributions, ethicality and efficiency conflict significantly: Self-nudging underperformed in efficiency terms, despite its favorable ethics and moderately cooperative contribution defaults.

The social nudge delivered significantly higher overall efficiency. When defaults can be made public, the conflict somewhat fades statistically: Active contributions were higher, on average, than under full contribution default. This effect, which offsets the "mechanical effect" of defaults, appears to be driven by subjects' beliefs. Still, changing the information structure does not alter the fundamental problem: In our experiment, the efficiency of self-nudging under public defaults stayed significantly below that of social nudging under (less demanding) private defaults. Not considered in this assessment are possible psychological cost on subjects from having to set a default.

In sum, choice architecture in social dilemmas remains particularly interesting and relevant in the light of current societal challenges. Discussions about how to reconcile divergent alignments between policy-makers and those targeted by their policies will continue and are vital in liberal democracies. Our results show that there are real and consequential impacts on the ethicality and efficiency of social dilemmas.

References

Ahn, T.-K., Isaac, R. M., and Salmon, T. C. (2008). Endogenous group formation. *Journal of Public Economic Theory*, 10(2):171–194.

Altmann, S. and Falk, A. (2009). The impact of cooperation defaults on voluntary contributions to public goods. Technical report, Unpublished Manuscript.

Altmann, S., Falk, A., and Grunewald, A. (2013). Incentives and information as driving forces of default effects. Technical report, IZA Discussion Paper No. 7610.

Angelovski, A., Di Cagno, D., Güth, W., Marazzi, F., and Panaccione, L. (2018). Does heterogeneity spoil the basket? the role of productivity and feedback information on public good provision. *Journal of Behavioral and Experimental Economics*, 77:40–49.

Arad, A. and Rubinstein, A. (2018). The people's perspective on libertarian-paternalistic policies. *The Journal of Law and Economics*, 61(2):311–333.

Arechar, A. A. (2018). Conducting interactive experiments online. *Experimental Echarneconomics*, 1:99–131.

Arechar, A. A., Gächter, S., and Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.

Banerjee, S. and John, P. (2021). Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, pages 1–16.

Blumenstock, J., Callen, M., and Ghani, T. (2018). Why do defaults affect behavior? experimental evidence from Afghanistan. *American Economic Review*, 108(10):2868–2901.

Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., Chmura, T. G., Cornelissen, G., Døssing, F. S., Espín, A. M., et al. (2017). Registered replication report: Rand, greene, and nowak (2012). *Perspectives on Psychological Science*, 12(3):527–542.

Briscese, G. (2019). Generous by default: A field experiment on designing defaults that align with past behaviour on charitable giving. *Journal of Economic Psychology*, 74:102187.

Bruns, H. and Perino, G. (2021). Point at, nudge, or push private provision of a public good? *Economic Inquiry*, 59:996–1007.

Bucher, T., Collins, C., Rollo, M. E., McCaffrey, T. A., De Vlieger, N., Van der Bend, D., Truby, H., and Perez-Cueto, F. J. (2016). Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. *British Journal of Nutrition*, 115(12):2252–2263.

Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P. (2019). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Cason, T. N., Saijo, T., Yamato, T., and Yokotani, K. (2004). Non-excludable public good experiments. *Games and Economic Behavior*, 49(1):81–102.

Charness, G., Gneezy, U., and Rasocha, V. (2021). Experimental methods: Eliciting beliefs. *Journal of Economic Behavior & Organization*, 189:234–256.

Dawes, R. M. and Thaler, R. H. (1988). Anomalies: cooperation. *Journal of Economic Perspectives*, 2(3):187–197.

Diederich, J., Goeschl, T., and Waichman, I. (2016). Group size and the (in)efficiency of pure public good provision. *European Economic Review*, 85:272–287.

Engel, C. and Kurschilgen, M. (2020). The fragility of a nudge: the power of self-set norms to contain a social dilemma. *Journal of Economic Psychology*, 81:102293.

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

Feito-Ruiz, I., Renneboog, L., and Vansteenkiste, C. (2020). Elective stock and scrip dividends. *Journal of Corporate Finance*, 64:101660.

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–56.

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

Fosgaard, T. R. and Piovesan, M. (2015). Nudge for (the public) good: how defaults can affect cooperation. *PloS one*, 10(12):e0145488.

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

Goeschl, T. and Lohse, J. (2018). Cooperation in public good games. calculated or confused? *European Economic Review*, 107:185–203.

Hagman, W., Andersson, D., Vastfjall, D., and Tinghog, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, 6(3):439–453.

Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. Random House.

Hokamp, E. G. and Weimann, J. (2021). Nudging openly–an experimental analysis of nudge transparency in a public goods setting. *German Economic Review*.

Horton, J. J. (2011). The condition of the turking class: Are online employers fair and honest? *Economics Letters*, 111(1):10–12.

Isaac, R. M. and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199.

Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group size and the voluntary provision of public goods. *Journal of Public Economics*, 54:1–36.

Johnson, E. J. and Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649):1338–1339.

Kaiser, M., Bernauer, M., Sunstein, C. R., and Reisch, L. A. (2020). The power of green defaults: the impact of regional variation of opt-out tariffs on green energy demand in germany. *Ecological Economics*, 174:106685.

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

Kimbrough, E. O. and Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3):608–638.

Madrian, B. C. and Shea, D. F. (2001). The power of suggestion: Inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, 116(4):1149–1187.

Mills, S. (2020). Personalized nudging. *Behavioural Public Policy*, pages 1–10.

Nagatsu, M. (2015). Social nudges: their mechanisms and justification. *Review of Philosophy and Psychology*, 6(3):481–494.

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

Normann, H.-T., Requate, T., and Waichman, I. (2014). Do short-term laboratory experiments provide valid descriptions of long-term economic interactions? a study of cournot markets. *Experimental Economics*, 17(3):371–390.

Osborne, M. J., Rosenthal, J. S., and Turner, M. A. (2000). Meetings with costly participation. *American Economic Review*, 90(4):927–943.

Pecorino, P. and Temimi, A. (2007). Public good provision in a repeated game: The role of small fixed costs of participation. *Public Choice*, 130(3-4):337–346.

Reijula, S. and Hertwig, R. (2020). Self-nudging and the citizen choice architect. *Behavioural Public Policy*, pages 1–31.

Schelling, T. C. (1978). Egonomics, or the art of self-management. *The American Economic Review*, 68(2):290–294.

Shank, D. B. (2016). Using crowdsourcing websites for sociological research: The case of amazon mechanical turk. *The American Sociologist*, 47(1):47–55.

Sunstein, C. R. (2013). Impersonal default rules vs. active choices vs. personalized default rules: A triptych. *Active Choices vs. Personalized Default Rules: A Triptych (May 19, 2013)*.

Sunstein, C. R. (2015). Nudges do not undermine human agency: A note. *Journal of Consumer Policy*, 38(3):207–210.

Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, 1(1):4–25.

Sunstein, C. R. and Reisch, L. A. (2013). Green by default. *Kyklos*, 66:398–402.

Tannenbaum, D., Fox, C. R., and Rogers, T. (2017). On the misplaced politics of behavioural policy interventions. *Nature Human Behavior*, 1:0130.

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1):S164–S187.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, And Happiness*. Yale University Press.

Thaler, R. H. and Tucker, W. (2013). Smarter information, smarter consumers. *Harvard Business Review*, 91(1):44–54.

Thunström, L., Gilbert, B., and Ritten, C. J. (2018). Nudges that hurt those already hurting–distributional and unintended effects of salience nudges. *Journal of Economic Behavior & Organization*, 153:267–282.

Tontrup, S. and Sprigman, C. J. (2019). Experiments on self-nudging, autonomy, and the prospect of behavioral-self-management. Technical Report 19-41, NYU Law and Economics Research Paper.

Yeung, K. (2017). 'hypernudge': Big data as a mode of regulation by design. *Information, Communication & Society*, 20(1):118–136.

Appendix

A Online Supplementary Material

A.1 Selection and Sample Characteristics

When signing up for the experiment with the Internet polling company, subjects had to enter a selection of sociodemographic data. This allows us to test for possible selection effects at the point at which subjects transfer from the registration stage to round 1 of the experiment. In other words, we can test whether there are indication that registered subjects with certain characteristics were significantly more likely not to 'show up' for round 1. Table A.1 reports the results of a logit regression that the propensity to not show up for round 1 is slightly elevated for females (their probability of non-shows is 6 percent higher than of males). Non shows are also reduced for the income group between 900-1200 EURO (but we do not find consistent effect of income on dropping out).

Table A.1: Selection analysis: Logistic regression

	Dropped out
Female	0.303***
	(0.116)
Age: 25-34	0.232
	(1.236)
Age: 35-44	-0.110
	(1.235)
Age: 45-54	-0.293
	(1.234)
Age: 55-64	-0.293
	(1.235)
East Germany	-0.140
	(0.145)
Berlin	0.162
	(0.258)
Academic education	-0.112
	(0.126)
Income: 900-1200 EUR	-0.825**
	(0.333)
Income: 1300-1500 EUR	0.059
	(0.267)
Income: 1500-2000 EUR	0.002
	(0.262)
Income: 2000-2600 EUR	-0.307
	(0.246)
Income: 2600-3600 EUR	-0.253
	(0.246)
Income: 3600-5000 EUR	-0.535**
	(0.253)
Income: above 5000 EUR	-0.382
	(0.329)
Constant	-0.611
	(1.248)
Observations	1532
Pseudo-R^2	0.025

Logit model estimation output- There were only 2 (4) potential participants below the age of 18 (above the age of 65). Income baseline category: Below 900 EUR per month. "Academic education" includes subjects who studies or are studying for an academic degree. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

Table A.2: Sample characteristics (percentage)

Female	49.72
Age: Below 18	0.19
Age: 18-24	6.30
Age: 25-34	17.13
Age: 35-44	23.43
Age; 45-54	29.07
Age: 55-64	23.61
Age: 65 and older	0.28
Academic education (total)	34.64
Apprenticeship / vocational training	46.02
Professional school / tertiary college	5.19
Master of crafts / technician	4.91
College degree	13.06
University degree	17.69
Other occupational degree	0.93
No apprenticeship or degree	3.52
Student at college or university	3.15
In apprenticeship or vocational training	3.80
Student in school	1.39
Ph.D. student	0.37
Residence: West Germany	74.54
Residence: East Germany	20.28
Residence: Berlin	5.19
Income: below 900 EUR	6.72
Income: 900-1200 EUR	7.3
Income: 1300-1500 EUR	9.06
Income: 1500-2000 EUR	10.52
Income: 2000-2600 EUR	20.06
Income: 2600-3600 EUR	19.57
Income: 3600-5000 EUR	20.74
Income: Above 5000 EUR	6.04

Notes: "Academic education (total)" includes subjects who studies or are studying for an academic degree

Table A.3: Sample characteristics by treatment

	Female (%)	Age (C)	Academic (%)	Income (C)
SELF PuD-PuC	0.47	4.47	0.43	5.25
	(0.05)	(0.11)	(0.05)	(0.18)
SELF PuD-PvC	0.53	4.62	0.29	5.20
	(0.05)	(0.10)	(0.04)	(0.18)
SELF PvD-PuC	0.56	4.17	0.30	5.12
	(0.05)	(0.12)	(0.04)	(0.19)
SELF PvD-PvC	0.44	4.49	0.36	4.71
	(0.05)	(0.11)	(0.05)	(0.21)
SELFISH PuD-PuC	0.57	4.46	0.31	5.21
	(0.06)	(0.14)	(0.05)	(0.19)
SELFISH PuD-PvC	0.55	4.38	0.32	5.07
	(0.06)	(0.14)	(0.05)	(0.20)
SELFISH PvD-PuC	0.44	4.46	0.30	4.41
	(0.06)	(0.15)	(0.06)	(0.24)
SELFISH PvD-PvC	0.53	4.50	0.42	5.08
	(0.06)	(0.15)	(0.06)	(0.23)
SOCIAL PuD-PuC	0.45	4.59	0.38	5.16
	(0.06)	(0.15)	(0.05)	(0.20)
SOCIAL PuD-PvC	0.44	4.87	0.35	5.16
	(0.05)	(0.13)	(0.05)	(0.24)
SOCIAL PvD-PuC	0.49	4.31	0.32	5.04
	(0.06)	(0.14)	(0.05)	(0.24)
SOCIAL PvD-PvC	0.49	4.33	0.41	4.69
	(0.06)	(0.14)	(0.06)	(0.23)
F-test	F (11, 1068):	F (11, 1068):	F (11, 1053):	F (11, 1015):
P-value	p=0.506	p=0.029	p=0.479	p=0.160

Note: Distribution of demographics by treatment (means and std. errs in parenthesis). "Age (C)" is age category, "Academic (%)" includes subjects who studies or are studying for an academic degree, Income (C) is income category. Age and income categories are displayed in Table A.2. F-tests are result of a joint-significance F-test in a linear regression where depended variable (female, age, academic, income) is regressed on dummies of the treatments.

Table A.4: Average total and active contributions, and non-participation

	Total Cont. (1)	Active Contr. (2)	No participation (3)
SELF PuD-PuC	34.82	34.65	0.07
	(2.04)	(2.05)	(0.01)
SELF PuD-PvC	35.88	35.72	0.09
	(1.79)	(1.72)	(0.02)
SELF PvD-PuC	31.37	30.90	0.13
	(1.74)	(1.78)	(0.02)
SELF PvD-PvC	34.64	34.69	0.12
	(2.09)	(2.28)	(0.02)
SELFISH PuD-PuC	33.08	35.78	0.09
	(2.74)	(2.65)	(0.02)
SELFISH PuD-PvC	26.25	29.96	0.13
	(2.17)	(2.60)	(0.03)
SELFISH PvD-PuC	32.37	35.84	0.11
	(2.57)	(2.69)	(0.02)
SELFISH PvD-PvC	30.05	33.19	0.11
	(3.00)	(3.02)	(0.03)
Social PuD-PuC	38.99	35.63	0.09
	(2.88)	(2.86)	(0.02)
Social PuD-PvC	38.71	34.50	0.10
	(2.18)	(2.30)	(0.02)
Social PvD-PuC	42.18	36.75	0.15
	(2.53)	(2.23)	(0.03)
Social PvD-PvC	41.53	38.21	0.10
	(2.48)	(2.22)	(0.03)

Note: Average total contribution, active contributions, and non participation over the 10 rounds (the data was first averaged per group).

B Robustness: Additional Regressions

Table B.1: Robustness: OLS regressions with two-way clustered std. err (PvD-PuC)

	Total contributions (1)	Active contributions (2)	No participation (3)
SOCIAL	9.595**	1.354	0.027
	(3.790)	(3.426)	(0.032)
SELF	-0.614	-4.039	0.020
	(3.325)	(3.219)	(0.035)
Round (mean centered)	-0.988**	-0.918***	-0.004**
	(0.353)	(0.262)	(0.001)
SOCIAL × Round	1.428**	1.082**	0.002
	(0.624)	(0.337)	(0.004)
SELF × Round	0.299	0.094	0.006*
	(0.304)	(0.324)	(0.003)
Constant	36.544***	41.025***	0.132*
	(5.133)	(4.546)	(0.065)
Additional Control	Yes	Yes	Yes
Observations	2470	2209	2223
R-squared	0.073	0.057	0.053

Note: Estimations with two-way robust clustered standard errors (across (i) groups and (ii) rounds). Baseline is the SELFISH default condition. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. Additional controls include gender, age, region of residence, education, income, weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. Model 3 is based on a linear probability model. *, **, and *** denote significance level of at least 10%, 5%, and 1%, using two-sided tests throughout.

Table B.2: Random-effects GLS regressions of total contributions, active contributions, and participation

	Total contributions (1)	Active contributions (2)	No participation (3)
		PuD-PuC	
SOCIAL	5.546	-0.505	-0.001
	(3.920)	(3.613)	(0.027)
SELF	1.165	-1.658	-0.001
	(3.147)	(3.130)	(0.022)
Round (mean centered)	-1.423***	-1.205***	0.005
	(0.306)	(0.254)	(0.003)
SOCIAL × Round	0.807	0.661	-0.014**
	(0.536)	(0.467)	(0.005)
SELF × Round	0.890**	0.587*	-0.003
	(0.374)	(0.339)	(0.004)
Constant	32.230***	37.618***	0.069
	(4.614)	(4.501)	(0.044)
Observations	2640	2451	2376
R-squared (overall)	0.08	0.08	0.06
		PuD-PvC	
SOCIAL	8.639***	1.843	-0.017
	(3.242)	(3.403)	(0.032)
SELF	8.441***	4.727	-0.031
	(3.039)	(3.192)	(0.032)
Round (mean centered)	-0.859***	-0.546***	0.007
	(0.202)	(0.196)	(0.005)
SOCIAL × Round	0.275	-0.282	-0.009*
	(0.359)	(0.342)	(0.005)
SELF × Round	0.128	-0.207	-0.002
	(0.278)	(0.292)	(0.006)
Constant	30.277***	34.906***	0.154***
Observations	2550	2310	2295
R-squared (overall)	0.11	0.08	0.09
		PvD-PvC	
SOCIAL	11.678***	5.105	0.012
	(3.805)	(3.718)	(0.034)
SELF	3.918	1.020	0.030
	(3.556)	(3.718)	(0.031)
Round (mean centered)	-0.812***	-0.361	0.006**
	(0.237)	(0.262)	(0.003)
SOCIAL × Round	0.770**	-0.035	-0.005
	(0.389)	(0.364)	(0.005)
SELF × Round	0.043	-0.481	-0.002
	(0.275)	(0.304)	(0.004)
Constant	28.969***	33.086***	0.068
	(5.468)	(5.673)	(0.047)
Observations	2460	2240	2214
R-squared (overall)	0.07	0.06	0.05

Note: Robust standard errors are clustered for experimental groups. Baseline is the SELFISH default condition. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. Additional controls are included all estimators, these including gender, age, region of residence, education, income, weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. Model 3 is based on a linear probability model.

Table B.3: Robustness: OLS regressions with two-way clustered std. err. of total contributions, active contributions, and participation

	Total contributions (1)	Active contributions (2)	No participation (3)
SOCIAL	7.979**	0.589	0.041
	(3.354)	(3.152)	(0.029)
SELF	-1.279	-4.453	0.035
	(2.779)	(2.808)	(0.028)
Public Default Information (PuD)	-1.329	-1.075	0.020
	(2.651)	(2.783)	(0.023)
Private Contributions Information (PvC)	-4.058	-4.026	0.024
	(2.664)	(2.816)	(0.025)
SOCIAL × PuD	-1.849	-0.957	-0.041
	(3.665)	(3.696)	(0.032)
SOCIAL × PvC	4.029	4.185	-0.028
	(3.704)	(3.773)	(0.032)
SELF × PuD	3.040	2.999	-0.049
	(3.268)	(3.406)	(0.029)
SELF × PvC	6.375*	6.391*	-0.017
	(3.316)	(3.478)	(0.031)
Round (mean centered)	-1.040***	-0.794***	0.004***
	(0.162)	(0.044)	(0.001)
SOCIAL × Round	0.838*	0.366***	-0.006**
	(0.395)	(0.097)	(0.002)
SELF × Round	0.360***	0.049	-0.001
	(0.107)	(0.055)	(0.001)
Constant	34.775***	39.101***	0.090**
	(3.110)	(3.095)	(0.028)
Additional controls	Yes	Yes	Yes
Observations	10120	9210	9108
R-squared	0.052	0.029	0.039

Note: Estimations with two-way robust clustered standard errors (across (i) groups and (ii) rounds). Baseline is the SELFISH default condition. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. "Public Defaults (PuD)" means that information about average defaults is revealed to the group. "Private Contributions (PvC) " means that contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. Model 3 is based on a linear probability model. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

Table B.4: Robustness: OLS regressions with two-way clustered std. err. of beliefs variables

	Beliefs about nonpart. rate (1)	Beliefs about rate pos.ctr. (2)	Beliefs about pos. contrib. (3)	Beliefs about active contrib. (4)
SOCIAL	-0.014	-0.001	3.337	3.379
	(0.033)	(0.033)	(2.585)	(2.623)
SELF	-0.044	-0.004	-1.491	-1.637
	(0.027)	(0.027)	(2.172)	(2.206)
Public Defaults (PuD)	-0.035	0.009	2.923	2.736
	(0.023)	(0.024)	(1.994)	(1.983)
Private Contributions (PvC)	0.062**	-0.127***	-5.706**	-6.268**
	(0.023)	(0.025)	(2.013)	(1.985)
SOCIAL × PuD	0.019	-0.042	-5.545*	-6.042*
	(0.034)	(0.037)	(2.840)	(2.888)
SOCIAL × PvC	0.007	0.044	2.878	2.847
	(0.035)	(0.034)	(2.842)	(2.895)
SELF × PuD	-0.011	0.046	1.462	1.933
	(0.030)	(0.031)	(2.684)	(2.683)
SELF × PvC	0.034	0.049	6.543**	6.741**
	(0.028)	(0.030)	(2.675)	(2.686)
Round (mean centered)	-0.002	-0.006	-0.366***	-0.311***
	(0.002)	(0.005)	(0.047)	(0.047)
SOCIAL × Round	0.002	-0.002	0.471***	0.505***
	(0.002)	(0.002)	(0.112)	(0.132)
SELF × Round	0.001	0.002	0.334***	0.349***
	(0.002)	(0.002)	(0.087)	(0.081)
Constant	0.222***	0.817***	31.569***	29.945***
	(0.031)	(0.044)	(2.268)	(2.354)
Additional controls	Yes	Yes	Yes	
Observations	8135	8876	8647	8647
R-squared	0.052	0.068	0.050	0.053

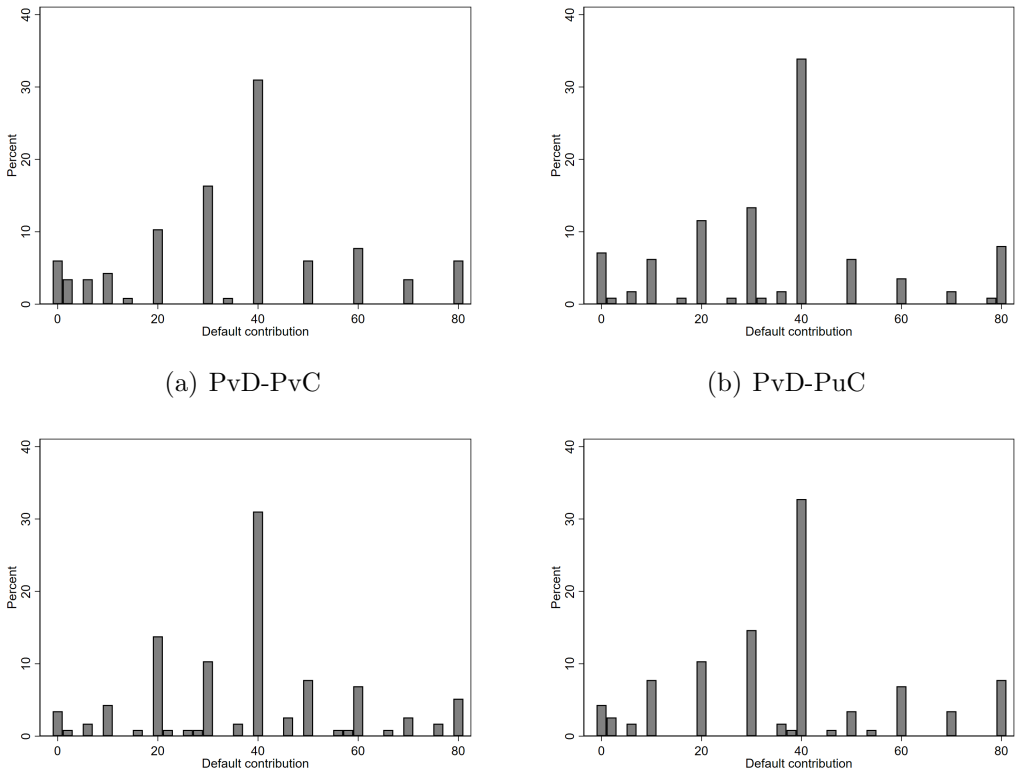
Estimations with two-way robust clustered standard errors (across (i) groups and (ii) rounds). Baseline is the SELFISH default rule. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. "Public Defaults (PuD)" means that information about average defaults is revealed to the group. "Private Contributions (PvC) " means that contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, and the weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

Table B.5: Robustness: OLS regressions with two-way clustered std. err. of total contributions and active contributions

	Total contributions (1)	Active contributions (2)
Beliefs about active contrib.	0.680***	0.681***
	(0.025)	(0.025)
SOCIAL	-1.806	-1.833
	(1.988)	(1.987)
SELF	-3.245	-3.262
	(1.896)	(1.895)
Public Defaults (PuD)	-2.474	-2.433
	(1.826)	(1.833)
Private Contributions (PvC)	0.990	0.952
	(1.943)	(1.950)
SOCIAL × PuD	2.917	2.879
	(2.359)	(2.362)
SOCIAL × PvC	1.137	1.185
	(2.357)	(2.365)
SELF × PuD	1.551	1.507
	(2.101)	(2.107)
SELF × PvC	1.475	1.515
	(2.227)	(2.233)
Round (mean centered)	-0.627***	-0.636***
	(0.055)	(0.052)
SOCIAL × Round	0.093	0.102
	(0.136)	(0.132)
SELF × Round	-0.093	-0.083
	(0.093)	(0.094)
Constant	19.905***	19.940***
	(2.432)	(2.425)
Additional controls	Yes	Yes
Observations	8647	8646
R-squared	0.319	0.320

Estimations with two-way robust clustered standard errors (across (i) groups and (ii) rounds).Baseline is the SELFISH default rule. "SELFISH" ("SOCIAL") denotes a treatment with an exogenously determined zero (full) contribution to the public good in case of non participation, whereas 'SELF" denote a treatment with self-determined default contributions. "Beliefs about active contrib" is constructed from belief elicitation items 2 and 3, indicating the average contribution of active participators. "Public Defaults (PuD)" means that information about average defaults is revealed to the group. "Private Contributions (PvC) " means that contributions are not revealed to the group. Additional controls include gender, age, region of residence, education, income, and the weekday of the experiment. The Round variable is mean-centered, hence, coefficient estimates correspond to marginal effects estimated at mean experimental round. All reported tests based on this regression are two-sided. *, **, and *** denote significance level of at least 10%, 5%, and 1%.

C Additional Figures



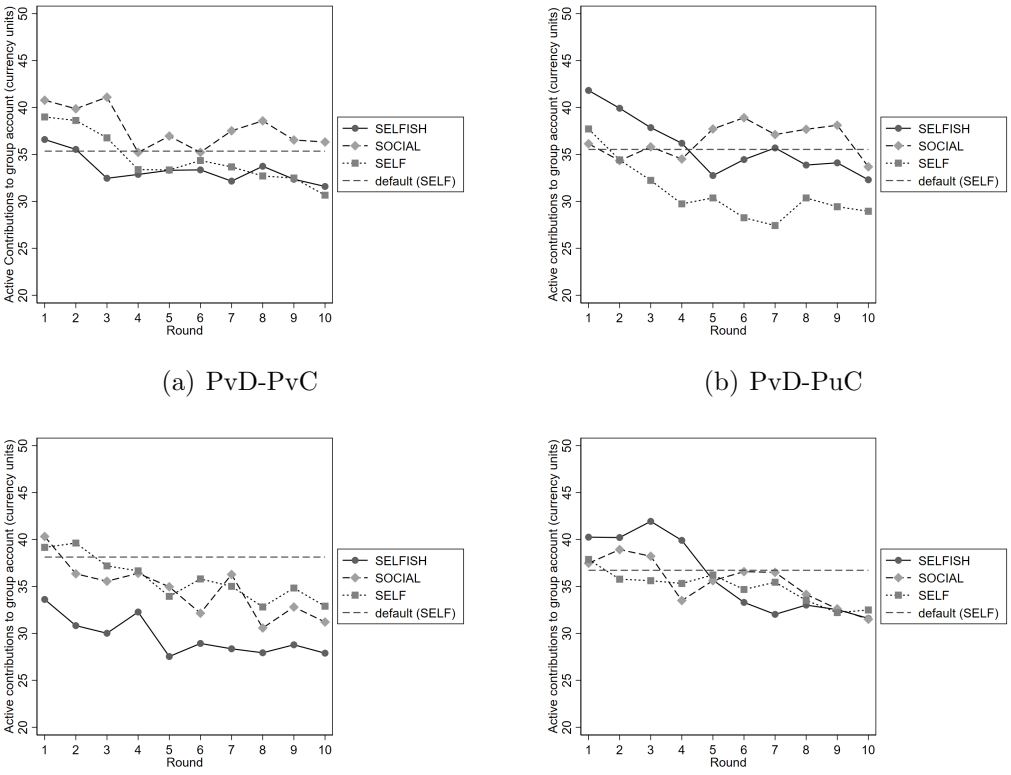
(a) PvD-PvC

(b) PvD-PuC

(c) PuD-PvC

(d) PuD-PuC

Figure C.1: Histograms of chosen non-participation default contribution values in SELF, by feedback information condition



(a) PvD-PvC

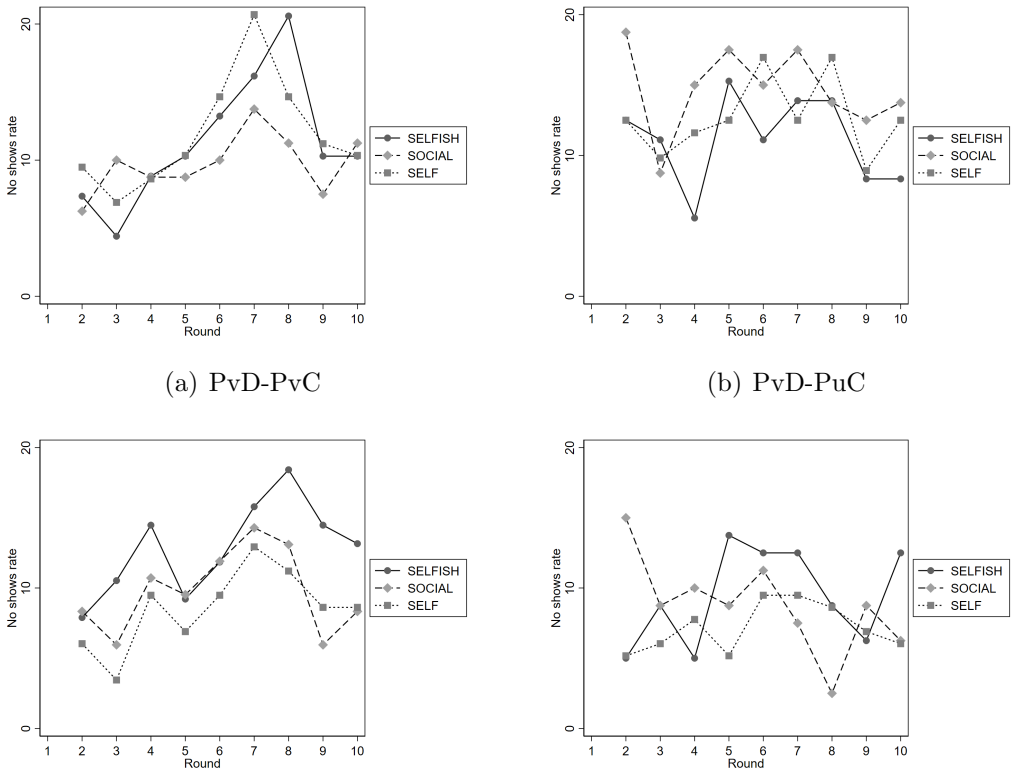
(b) PvD-PuC

(c) PuD-PvC

(d) PuD-PuC

Note: Horizontal dashed lines denote the average self-determined default contribution in SELF.

Figure C.2: Evolution of average active contributions



(a) PvD-PvC

(b) PvD-PuC

(c) PuD-PvC

(d) PuD-PuC

Figure C.3: Evolution of non-participation rates