Durham University Business School

Department of Economics

The Limits of Moral Wiggle Room: Plausible Deniability in Public Good Games

Danae Arroyos-Calvera & Andis Sofianos

Working Paper No. 3, 2026

Durham Business School working papers series ISSN 1749-3641

The Limits of Moral Wiggle Room: Plausible Deniability in Public Good Games

Danae Arroyos-Calvera[*] Andis Sofianos[†]

March 31st, 2026

Abstract

Although in the lab we often assume a deterministic relationship between contributions to public goods and their returns, in real life this relationship is often uncertain: consider corruption or inefficiencies in governments, or areas where our knowledge is limited such as climate change mitigation. Rational individuals may change their contributions to respond to this uncertainty, but we may also observe a reduction in contributions if people take advantage of the plausible deniability about one's actions that this uncertainty creates. The reduction of prosocial behaviour in the presence of plausible deniability was documented in Dana et al. (2007) and has since been widely replicated in two-person non-strategic settings. We test whether people exploit uncertainty in public good games (PGG) by using it as an excuse to free-ride. In two online studies that entail 5 between-subjects treatments, we varied whether the marginal per capita returns (MPCR) were certain or uncertain, and whether this uncertainty could give participants plausible deniability for their low or zero contributions. We found no change in free-riding or average contributions across treatments, suggesting that cooperation under uncertainty is accurately captured by the standard PGGs with certain MPCR, and that uncertainty need not hamper cooperation in social dilemmas.

Keywords: Moral wiggle room, Plausible Deniability, Public Good Game, Experiment

[*]University of Birmingham, UK; d.arroyoscalvera@bham.ac.uk (Corresponding author)
[†]Durham University, UK; andis.sofianos@durham.ac.uk

Acknowledgements: We thank Christoph K. Becker, Jordi Brandts, Simon Gaechter, Ben Grodeck, Zachary Grossman, Marco Lambrecht, David Levine, Johannes Lohse, Joerg Oechssler, Joël van der Weele; the Strategic Information Network members Despoina Alempaki, Valeria Burdea, Andrea Isoni, Eugene Malthouse, Rebecca McDonald, Hamideh Mohtashami Borzadaran, and Daniel Read; and the seminar and conference participants at Birmingham, Lancaster, Loughborough and Newcastle Universities, Economic Science Association meetings, the 2022 International Conference on Social Dilemmas, International Meetings on Experimental and Behavioural Social Sciences, Birmingham Behavioural and Experimental Economics workshops, the University of Birmingham - NYU Abu Dhabi workshop in Behavioural and Experimental Economics, the 1st Women in Behavioural and Experimental Economics workshop, and the 17th Nordic Conference in Behavioral and Experimental Economics for their constructive comments and suggestions on this paper. The funding provided by the University of Heidelberg, University of Birmingham, and Durham University is gratefully acknowledged. Ethics approval was obtained from the German Association for Experimental Economic Research (number DVtrkfik) and the University of Birmingham ethics board (ERN_20-1020).

Department Economics
Durham University Business School
Mill Hill Lane
Durham DH1 3LB, UK
Tel: +44 (0)191 3345200
https://www.durham.ac.uk/business/about/departments/economics/

1 Introduction

Governments and charities around the world rely on citizens' contributions to provide public goods such as healthcare, road safety, nature conservation, and clean air. The non-excludable nature of these goods makes them prone to the free-rider problem, as once these goods are into place everyone can benefit from them, regardless of whether they have contributed towards them or not. The need to better understand how to achieve and sustain cooperation has lead to the emergence of a large experimental literature that uses public good games (PGG).

In this paper, we are particularly interested in situations where contributions may not translate into the provision of the public good. This could be due to multiple reasons. For example, government inefficiencies, which according to a former UK Treasury minister cost the UK £35bn per year (BBC, 2022); fraud, bribery, and corruption, which the UK's National Health Service Counter Fraud Authority estimates cost £1.3bn per year (NHS Counter Fraud Authority, 2024); or simply, inherent uncertainty about the benefits of the provision for policies that entail a more long-term outlook, as in the case of climate change mitigation. The key commonality between all of these examples is that the link between contributions and the public good provision is uncertain, transforming the decision of whether to contribute and how much into a risky[1] choice problem.[2]

The uncertain link between contributions and public good provision may lead individuals to change their behaviour, compared to what they would do under certainty. Levati et al. (2009) theorise that, with risk, contributions in public good games will be lower, whilst Björk et al. (2016)'s model anticipates either no change or an increase in contributions. Experimental work featuring uncertain homogeneous returns[3] also presents mixed findings. Levati et al. (2009), Cárdenas et al. (2017), and Freundt and Lange (2021) report lower contributions under uncertainty. Meanwhile, Dickinson (1998) finds lower contributions, but only in the first four rounds and not statistically significant at the 5% significance level. Levati and Morone (2013) report no difference when they consider all rounds. Brennan et al. (2008), Stoddard et al. (2015), Björk et al. (2016) and Théroude and Zylbersztejn (2020) also find no difference in contributions across uncertainty treatments. In the words of Théroude and Zylbersztejn (2020), "this literature is quite small and disperse [...] and [...] the existing findings are quite mixed."

Risk has also been found to affect behaviour by obscuring the link between one's choices and their consequences on others. The presence of risk gives people plausible deniability with regards to whether the socially suboptimal outcome (if any) was brought by them being selfish or simply unlucky. This allows people to behave selfishly without having to incur the costs of this behaviour in terms of worsened image. Dana et al. (2007) coined the term "moral wiggle room" to refer to this plausible deniability when they found that people were less generous if the link between their

[1] The terms "risk" and "uncertainty" are used interchangeably to refer to a situation of multiple possible outcomes with known probabilities.

[2] Existing work has been conducted in the settings described in our examples (for example, Cagala et al., 2019; Drouvelis et al., 2025, who investigate cooperation in the presence of rent extraction). Our design does not map onto any of the above scenarios, which are only used to illustrate the pervasiveness of situations where contributions do not necessarily result in the provision of the public good. We instead abstract from the source of inefficiencies but focus on the mere presence of such frictions that could be acting as excuses for lower contributions.

[3] This implementation of uncertainty is the closest to our motivating examples, whilst still remaining tractable together with our experimental manipulations. There have been other implementations of risk in public good games, but these are not as relevant for our context. For example: uncertain endowments (Ebeling, 2013), uncertain implementation thresholds (Dannenberg et al., 2015), a probabilistic implementation of the public good (Dickinson, 1998; Gangadharan and Nemes, 2009), and uncertain individual returns (Fischbacher et al., 2014; Oechssler et al., 2022; Reischmann and Oechssler, 2018).

choice and the payoff to the other person was uncertain.[4] They found this in a binary dictator game and many studies since have replicated their findings in similar, non-strategic, two-person settings (e.g., Andreoni and Bernheim, 2009; Matthey and Regner, 2011; Feiler, 2014; Grossman, 2014; Grossman and Van Der Weele, 2017; Exley, 2016; Garcia et al., 2020; Brock et al., 2013). [5]

The plausible deniability that risk affords may reduce prosocial decisions in PGGs too. The *natural* uncertainty present in the risky variants of both dictator and public good games, which implies that it is not only up to the decision maker but also down to chance what the final payoff to all parties is, may weaken the link between each individual's giving decision and the benefit everyone receives. This would create moral wiggle room, which could be exploited by lowering contributions without the associated negative image cost. Public good games inherently feature *strategic* uncertainty as well: everyone's payoffs depend on one's own choices as well as those of others, but players cannot be certain of what others will choose when they are making their decision. As such, under strategic uncertainty, the link between each individual's contribution and the provision of the public good would become weaker. This could compound the effect of natural uncertainty and reduce contributions even further away from the social optimum.

In this paper, we investigate whether risk may be used as an excuse for less prosocial choices in public good games, taking contributions (further) away from the social optimum. To the best of our knowledge, uncertainty as a source of plausible deniability or *moral wiggle room*, has not yet been explored as a factor affecting contributions in public good games.

We answer this question using two online studies. In the first study, we implement a baseline treatment with certain returns, which aims to replicate the commonly found suboptimal contributions to the public good. We compare contributions in the baseline to those in a *Risk* treatment, with the goal of shedding more conclusive light on the effect of risk on contributions in PGGs. In the *Risk* treatment, similarly to other studies in the literature, group returns depend on the outcome of a coin toss but this does not offer plausible deniability towards others. In our *Plausible Deniability* (*PD*) treatment, we introduce a risk that each participant's contribution towards the public good is lost and participants appear as contributing nothing even when they intended to contribute a positive amount. This implies that a contribution of 0 cannot be attributed to a participant's selfishness with certainty: there is the possibility that, even when participants contributed some positive amount, they were unlucky and their contribution got lost. Thus, in this treatment, participants are granted plausible deniability about choosing to contribute nothing. By comparing the *Risk* and *PD* treatments, we can disentangle any effect that risk-induced plausible deniability may have on contributions and free-riding.

In the second study, we run a *No Plausible Deniability* (*NoPD*) treatment, which is identical to the *PD* treatment in terms of risk, but does not grant plausible deniability to free-riders, and a *Partial Plausible Deniability* (*PD-Partial*) treatment where a partial loss of the contribution (vs. the full loss in previous treatments) grants plausible deniability about lower contributions. We also administer the *PD* treatment again to serve as a benchmark for these two new treatments.

We find contributions do not significantly change under uncertainty, compared to the certain

[4] Dana et al. (2007) used "plausible deniability" to refer to the mechanism in their third study, where it was uncertain whether the participant had made the payoff-relevant decision or whether the computer had drawn the payoffs at random. They used "moral wiggle room" as the umbrella term for excuses that protect people's image from being tarnished when they are not behaving prosocially. In our two studies, the only source of plausible deniability, and therefore moral wiggle room, is our experimental manipulation. Thus, we use the two terms interchangeably.

[5] Examples of other sources of moral wiggle room that have been found to increase selfish behaviour, in strategic as well as non-strategic settings, include hidden information in a matching pennies game (Huang et al., 2024), a moral envelope game (Serra-Garcia and Szech, 2022), a trust game (Dufwenberg et al., 2001), and ultimatum games (e.g., Van Dijk et al., 2004; Guth et al., 1996), as well as anonymity in a variety of contexts (e.g., Christens et al., 2019; Dannenberg et al., 2022, 2023).

returns baseline. Most importantly, our participants did not free-ride more when uncertainty gave them plausible deniability; that is, they did not use uncertainty as an excuse to hide behind when making socially suboptimal contributions in the PGG. Equivalence tests provide statistical support to the absence of an effect of uncertainty and plausible deniability on contributions and free-riding. Our findings suggest that reducing uncertainty, and the plausible deniability that comes with it, would not increase individuals' cooperation in social dilemmas.

The rest of the paper is organised as follows. In the next section, we review the related literature and derive our predictions. In section 3, we outline our general experimental design and in sections 4 and 5 we describe the specific designs and results for study 1 and 2 respectively. In section 6, we discuss our findings and offer some concluding remarks.

2 Hypothesis: Risk, Other-Regarding Preferences, and Image

Previous work has theorised about the effect of risk, other-regarding preferences, and image concerns on voluntary contributions and, more generally, prosocial behaviour. In this section, we build upon that work to formulate our main hypothesis.

Levati et al. (2009) set up the decision of how much to contribute to a public good under uncertain marginal per capita returns (MPCR) as a risky gamble, where individuals would be equally happy to receive the certainty equivalent or the expected benefit from the contribution. Thus, the contribution that individuals choose is that which makes them indifferent between the certainty equivalent of the risky gamble and its expected benefit. The latter depends on their expectations about the MPCR and the contribution of others. The more risk averse an individual is, the lower the certainty equivalent will be, as they would happily leave money on the table if that means not being exposed to risk, while for risk seeking individuals, the opposite is true. A lower (higher) certainty equivalent amount would translate into a lower (higher) implied MPCR, which would hence induce lower (higher) contributions from risk averse (risk seeking) individuals.

These predictions change when other-regarding preferences are also considered. Björk et al. (2016) derive predictions that account for other-regarding preferences together with attitudes to risk. They start by setting up a model for a risk neutral individual where the weight of the other player's utility from the game payoffs is (linear and) captured by an unconditional altruism or warm glow parameter. They then show that, depending on how (constant relative) risk aversion is introduced in the utility function, contributions are predicted to either not change or increase. Namely, when risk aversion only affects the utility derived from one's own payoff but not that from others' payoffs, then contributions to the public good are predicted to increase with risk aversion. This result would be consistent with Brewer and Kramer (1986)'s psychological insight that stochastic environments create a "common fate group identity" which increases cooperation. In contrast, when risk aversion affects both components of the utility function, contributions are not predicted to change with one's attitude to risk.

Plausible deniability is hypothesised to affect people's other-regarding behaviour through an image channel (e.g., Grossman, 2014; Grossman and Van Der Weele, 2017). People derive utility from signalling, to themselves and to others, that they are good people (e.g., Andreoni and Bernheim, 2009; Bodner and Prelec, 2003; Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008). When making allocation decisions under certainty, individuals need to trade off the utility they will get from the material payoffs they will receive, with that from maintaining a good image. For example, when playing a dictator game, one can choose to keep the full £10 endowment and look like a selfish person; or to give away half of the endowment and be £5 worse off, but look like a good person. When risk offers plausible deniability, however, people can reap both the material and image payoffs as, even in situations when the decision maker secures most of the material benefit, this cannot be unequivocally attributed to their selfishness.

Dana et al. (2007) find that when people had plausible deniability to behave selfishly without having to update their image as good people, they did. On the basis of this, they attributed the observed generosity in the absence of plausible deniability to "an illusory preference for fairness", as if donations had been driven by social preferences, they would not have changed in the presence of plausible deniability. Grossman and Van Der Weele (2017) present a model that includes both social preferences and image: utility increases as a function of the material payoffs, which are weighted by a social preference parameter and one's image, which depends on whether the individual behaved selfishly or not, and whether information on whether they behaved selfishly is available to all parties.

Our research question is whether the plausible deniability that uncertainty creates may also increase self-serving behaviour in a public good game setting, compared to a situation without uncertainty. Closely related to this question, among other things, Flambard et al. (2024) hypothesise that individuals that exploit moral wiggle room would contribute less with uncertainty. However, the study does not include a plausible deniability manipulation that allows to directly test this. Here, we take this one step further, and experimentally disentangle the role of uncertainty from that of plausible deniability. To answer this question, we need to be able to disentangle the effect on contributions that stems from attitudes to risk (and any interactions with other-regarding preferences) from that of plausible deniability through the image channel. The existing heterogeneity concerning the effect of risk on contributions, both in terms of theoretical predictions as well as experimental results, calls for the need to disentangle the two using a single experimental paradigm.

Following from results in the moral wiggle room and signalling literature, we hypothesise that under plausible deniability, average contributions will decrease and, in the case where plausible deniability justifies free-riding (which we use to denote contributions of zero), free-riding will also increase.

3 Experimental Design

Participants in our study play a one-shot public good game that varies, between-subjects, along two dimensions: the uncertainty regarding the marginal per capita returns (MPCR), and whether participants have plausible deniability about their contributions. The pre-registered hypotheses and experimental design can be found on the following link: `https://www.socialscienceregistry.org/trials/6128`. Ethics approval was obtained from the German Association for Experimental Economic Research (number DVtrkfik) and the University of Birmingham ethics board (ERN_20-1020).

3.1 Public Good Game

The game is a standard linear public good game with group size of 4. Each player starts with an endowment of 10 Taler (each Taler is worth £0.10) and they have to decide how much of it to invest into the group's common project (c_i) and how much of it to keep for themselves (10-c_i). The contribution to the public good of each player is disclosed at the end of the game. Each player's payoff π_i is determined by equation (1): players receive the Taler they keep for themselves plus the returns of the common project, θ Taler for each Taler that the group invests into the common project (i.e., θ is the MPCR).

$$\pi_i(c_i) = 10 - c_i + \theta \sum_{j=1}^{4} c_j \tag{1}$$

As long as $\frac{1}{4} < \theta < 1$ and players are rational, self-interested payoff maximisers, participants playing the game will face a social dilemma: whilst the social optimum is for all players to contribute their whole endowment into the common pool, the Nash equilibrium is for each player to keep their whole endowment to themselves and contribute nothing.

Our participants play a single round of the game. This is a key element of our design that helps us rule out any potential influence that group dynamics, reputation, or moral balancing may have on our results. There is abundant evidence in the literature documenting that a large proportion of individuals are conditional cooperators (some examples: Fischbacher et al., 2001; Fischbacher and Gächter, 2010; Thöni and Volk, 2018) and the literature consistently finds that contributions towards the public good generally follow a downward trend as participants increasingly learn to free-ride (e.g., Bigoni and Suetens, 2012; Burton-Chellew and West, 2021; Grandjean et al., 2022; Herrmann et al., 2008). Furthermore, as Lambrecht et al. (2024) argue, decisions made beyond the first period choice are always confounded with the experience from previous rounds.

Meanwhile, playing multiple rounds of the game could induce participants to change their contributions to manage their reputation as well (Fehr and Fischbacher, 2003), potentially further confounding the effect of our manipulation. Finally, moral balancing, the tendency to alternate between moral and immoral behaviour to maintain one's self-image (Nisan and Horenczyk, 1990; Ploner and Regner, 2013) has been documented in public good games (Seçilmiş, 2018) and could also confound our results.

Given these, it would be very difficult to unequivocally attribute any increase in free-riding behaviour observed over multiple rounds to plausible deniability. Whilst learning about how these factors would interact with our treatments would be interesting, we leave this for future research.

Because participants only play one round of the game, it is crucial that they learn about the mechanics of the game before playing it. To that end, we provide detailed instructions, examples, comprehension questions, and feedback. We also ask subjective understanding questions at the end of the study, which we use in our robustness checks. Throughout the study we administer attention checks which, unless answered correctly, would prevent participants from completing the study. These can be found in appendix A.

3.2 Other Measures

After playing the PGG, participants fill in questions about their age, gender, perceived financial status, familiarity with economics and/or game theory, trust in institutions (adapted from the Trust in Government Survey) and political views (captured by the standard question from the World Values Survey).

Two to six weeks after this, participants were invited to complete some additional tasks that were not identified as being part of the same study. In this questionnaire, we measure participants' risk preferences with a Holt and Laury (2002) multiple price list, and other-regarding preferences using Murphy et al. (2011)'s social value orientation questionnaire. The questionnaire was fully incentive-compatible, with one question, selected at random for each participant, being paid out.

We use the data from these questions as control variables in our regressions. The full set of instructions is reproduced in appendix A.

4 Study 1

4.1 Treatments

In the first study, we administer a *Baseline* treatment where the returns of the public good are certain, with $\theta = 0.4$, and two treatments with an uncertain θ equal to 0.4 in expectation.[6] The two treatments differ on whether participants can hide behind uncertainty to justify to their group members a non-contribution to the public good, which we refer to them as having plausible deniability.

The *Risk* treatment replicates the design of previous experiments testing the effect of uncertainty on contributions. The θ is equally likely to be 0.3 or 0.5, and therefore 0.4 in expectation. The realised θ is common for all group members. In this treatment, players have no plausible deniability for contributions of zero. Assuming risk neutrality, the standard equilibrium for one-shot public good games of free-riding and zero contributions is preserved. Naturally, with different risk preferences, other equilibria are possible; for this reason we elicit risk attitudes and control for these in the analysis of the data.

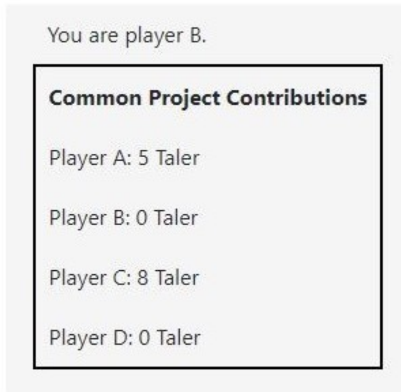
In the *Plausible Deniability* (*PD*) treatment, we set $\theta = 0.5$ and introduce a 20% chance that each player's contribution to the public good is lost before reaching the common project. This uncertainty implementation gives players plausible deniability, that is, the chance to justify their lack of contributions in front of others as bad luck rather than a lack of cooperation: when participants observe a contribution of zero from another player, they cannot infer whether that other player had indeed chosen to contribute nothing, or whether they had been unlucky and their contribution was lost.[7] If participants exploit this plausible deniability, we would expect an increase in free-riding, and therefore a decrease in average contributions. The chance that a player's contribution is lost is independent for each player, and that, together with the parameters above, imply an expected θ equal to 0.4, making this treatment directly comparable to the *Baseline* and *Risk* treatments.

Figure 1 displays an example of how contributions of individual group members are shown to all group members after decisions are made. Participants were informed that after each group member had decided how much they would like to contribute towards the public good, everyone's contributions would be displayed as in Figure 1. Specifically, while viewing Figure 1, the participants read about the decisions of an example group and were given this description: "*This is how the group contributions to the common project will be shown to you and your group members. Imagine you are player B.*" Notice that in the *Baseline* and the *Risk* treatments, player D's contribution of zero could be unequivocally attributed to player D not contributing anything to the common good and free-riding on the contributions of others. Instead, in the *PD* treatment, when observing that player D has contributed zero, other group members cannot tell for sure whether this is a free-riding attempt by player D or whether they were prosocial but unlucky, so whatever they had

[6] We also administer a baseline with $\theta = 0.5$. This additional baseline provided a comparison of choices with the same nominal MPCR (θ=0.5) as in the *PD* treatment, to check whether participants were behaving as if they were ignoring the exogenous risk that contributions would be lost. This baseline's MPCR also corresponds to the higher possible value for the MPCR in the *Risk* treatment, allowing us to check if participants in this treatment had behaved as if they were optimistic about the MPCR their group would get. Given the overall null effects we report when comparing behaviour across treatments, we omit any detailed analysis of the *Baseline*-05 treatment. We report the average contribution for this additional baseline treatment in the appendix for completeness.

[7] Our manipulation has a similar structure to that in earlier studies of the trust game (e.g., Charness and Dufwenberg, 2006; Regner, 2018; Regner and Matthey, 2021; Van der Weele et al., 2014) and the dictator game (*plausible deniability* treatment in Dana et al., 2007; and Andreoni and Bernheim, 2009) where nature's intervention may lead to an unfavourable outcome, and other players cannot observe whether nature has intervened. We are the first to apply such a manipulation in a public good game setting.

Figure 1: Summary of Contributions on Participant Screens.



contributed had been lost. This is the plausible deniability that free-riders enjoy in this treatment if they choose to contribute nothing to the public good.

We can identify the effect of uncertainty on contributions by comparing the *Baseline*, where there is no uncertainty involved and no plausible deniability granted to the participants, to the *Risk* treatment. By comparing the *Risk* treatment, where there was uncertainty but no plausible deniability, to the *PD* treatment, we can observe the effect of plausible deniability on contributions.

Implementation Details

We implemented our online study using oTree (Chen et al., 2016). We recruited participants using Prolific (`https://www.prolific.com`). The median participant took 10 minutes to complete the study, which results in an hourly pay of £14.

To ensure data quality, we only recruited participants with a Prolific approval rate of 95%+. We also included comprehension questions, provided feedback on their answers, and asked participants to report the level of their subjective understanding of the instructions, which they had the possibility to revisit. Lastly, we included attention checks, which prevented participants that failed two of them from completing the study and therefore be included in our analysis. We also asked participants to report whether they trusted the instructions, whether they completed the study on their own, and whether they were focused on the study while they were completing it which we use as robustness checks. These questions can be found in the instructions in appendix A.

We determined the required sample size using a power calculation, to ensure our study would have sufficient statistical power to detect an economically meaningful effect. We used the average contribution and standard deviation of existing similar work as input (Oechssler et al., 2022), and set a difference of one-third of a standard deviation across treatments as an economically meaningful effect. This corresponds to a Cohen's d = 0.33, which is considered a small-to-medium effect. We decided on this effect size by making a trade-off between the economic significance of the difference between treatments and the number of participants that would be required to observe this difference. The minimum required sample size to detect a difference of one third of a standard deviation is 112, which is why we aim for 120 participants per treatment.[8]

[8] To observe a difference of one half of a standard deviation, only 51 participants per treatment are required, whereas for a difference of one sixth of a standard deviation, a sample of 446 per treatment would be required.

Table 1: Participant Demographics and Other Characteristics – Study 1

	Baseline	*Risk*	*PD*
N	118	119	117
Age	37.64	39.97	38.24
Female	0.5	0.51	0.52
Econ Training	0.15	0.16	0.14
Trust in Judicial System	5.75	5.50	5.85
Trust in Government	3.33	3.49	3.86
Trust in Elections	5.63	5.86	6.08
Politically Left(1)-Right(10)	4.34	4.58	4.39
SVO: Competitive	0.02	0.03	0.02
SVO: Individualist	0	0	0
SVO: Prosocial	0.49	0.48	0.39
SVO: Altruistic	0.49	0.48	0.59
Risk Aversion	5.83	5.76	6.10

Notes: Econ Training refers to participants reporting that they have trained in economics. The SVO categories report the proportion of participants consistent with each classification. Risk aversion refers to number of safe choices made in the MPL task.

4.2 Results

In table 1, we report the characteristics of the participants in the *Baseline*, *Risk*, and *PD* treatments.[9] We find no significant differences in personal characteristics across treatments (all $p > 0.05$, exact Mann–Whitney U, MWU, test). We do not find significant differences either in terms of the fraction of correct answers in the comprehension questions, on self-reported understanding of the instructions, nor on self-reported attentiveness.[10] Practically all participants reported to trust the instructions to be accurate (98%), that they completed all their tasks on their own (100%), and were focused exclusively on the tasks while completing them (98%). Thus, we are confident that the allocation of participants into the treatments is sufficiently random and we do not face any selection issues.

Since we collected the risk preference and pro-sociality measures some weeks after the main study, it is important to detail any attrition in-between the two parts. Out of 354 participants across the three treatments, 302 returned and completed the second part, giving an attrition rate of roughly 15%. This is similar across the three treatments and given the random allocation across treatments this should not influence the results when comparing treatments.

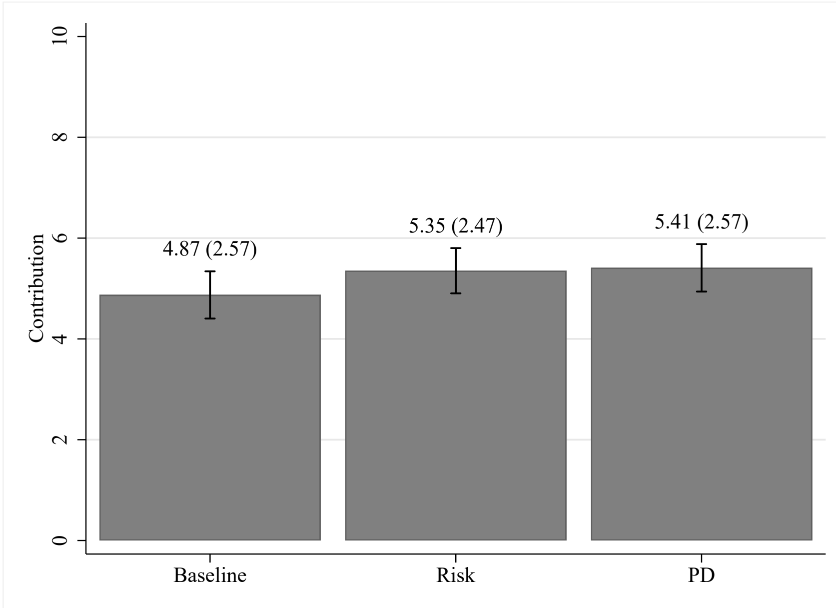
In figure 2, we present the average contributions to the public good across the three treatments.[11]

[9] We include all participants that made a decision on how much to contribute. As the experiment was implemented online, participants were matched into groups after they made their contribution decision so as to minimise mismatching and long waiting times. In very few instances, participants were not successfully matched into a group of four and, thus, they were not able to proceed beyond the contribution decision. For this reason, the number of participants in each treatment is not always divisible by four.

[10] The lowest average fraction of correct answers to the comprehension questions was in the *Baseline* treatment, with 74% correct answers. In the *Risk* treatment, the average fraction of correct answers is 79% and in the *PD* treatment, it is 84%. Pairwise comparisons reveal that there are no significant differences across treatments, with all $p > 0.09$.

[11] We report the full distribution of amounts sent, across all treatments, in figure A.2 in the appendix.

Figure 2: Average Contribution – Study 1.



Note: We denote the average *intended* contribution above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

The amount reported for the *PD* treatment reflects participants' *intended* contribution, that is, the amount participants choose to contribute towards the public good. This might differ from their *realised* contribution, as there was a chance that the amount they sent would be be lost. We analyse the former, and only use the latter to compute participant payments.

Contributions appear very similar across the three treatments. On average, participants in the *Baseline* sent 4.87 Taler, while those in the *Risk* treatment sent 5.35. These two values are not significantly different ($p = 0.1475$, exact MWU test). Thus, we find that, keeping the expected MPCR constant, the introduction of uncertainty does not have an effect on contributions. Comparing the *Risk* and *PD* treatments, we find that the average contributions are very similar, 5.35 in *Risk* and 5.41 in *PD*. We find no evidence of a significant difference in contributions between these two treatments either ($p = 0.8653$, exact MWU test).[12]

To confirm that the data supports the null effects that we are reporting, we perform equivalence tests. Using traditional significance tests, we can reject the absence of an effect but cannot confirm that the absence of an effect is statistically supported. Equivalence tests allow us to confirm if the data in question support a null effect by indicating whether the data between treatments are equivalent or not (Lakens, 2017). Concretely, we test if the experiment is powered enough to detect a difference between treatments of a magnitude of at least one third of a standard deviation (given the distribution of decisions across all three treatments), that is, an 'economically worthwhile' treatment difference which corresponds to a Cohen's d = 0.33, or a small-to-medium effect.[13]

An equivalence test for a difference of 1/3 of a standard deviation (0.86 Taler in our case) comes out inconclusive, but when we consider instead a difference of 1/2 of a standard deviations (1.3

[12]The comparison between the *Baseline* treatment and *PD* treatment is also not statistically significant ($p = 0.1132$, exact MWU test).

[13]More specifically, we use the two one-sided tests (TOST) procedure (Schuirmann, 1987) which works by specifying a lower and upper bound for a minimum 'worthwhile' treatment difference. A treatment difference falling within these bounds is deemed equivalent, and thus, an effect size not worthwhile to examine.

Taler), the test confirms the data from the two treatments are equivalent. This indicates that, if there is an undetectable difference between treatments, its magnitude would be smaller than 1.3 Taler.

We report regression analysis where we control for demographics, risk attitudes, and prosociality in table A.1 in the appendix.[14] Consistently with the analysis so far, we find that neither the *Risk* nor the *PD* treatments have a significant effect on contributions.

We summarise the findings from the analysis so far with:

Study 1 - Result 1. *Contributions towards the public good are unaffected by uncertainty.*

Study 1 - Result 2. *Plausible deniability about contributing nothing does not have a significant effect on contributions.*

The *PD* treatment manipulation provides participants with plausible deniability about free-riding (i.e., contributing zero). We hypothesised that plausible deniability would reduce contributions through increasing the number of people contributing nothing and hiding behind this plausible deniability. The lack of significant differences in average contributions might be masking an increase in free-riding from some participants, paired with other participants potentially increasing their contribution in anticipation of increased free-riding. It is therefore instructive to analyse the extent of free-riding behaviour across the treatments.

Figure 3 depicts the free-riding proportion in each of the three treatments. We find no difference in the prevalence of free-riding behaviour between the *Risk* and *PD* treatments ($p = 0.9766$, exact MWU test). Once again, as we are reporting a null effect, we perform equivalence tests for an effect equal to one third of a standard deviation. We find that the free-riding proportion is equivalent between *Risk* and *PD*. In table A.1 in the appendix, we report regression analysis where we control for various participant characteristics as described further above. The regressions corroborate the results from the statistical tests reported: free-riding behaviour is statistically indistinguishable between the *Risk* and *PD* treatments.[15] Incidentally, even though figure 3 shows an apparent drop in free-riding behaviour in the *Risk* and *PD* treatments in comparison to the *Baseline* this is not statistically significant (*Baseline* vs. *Risk*, $p = 0.1883$; *Baseline* vs. *PD*, $p = 0.3229$, exact MWU tests).

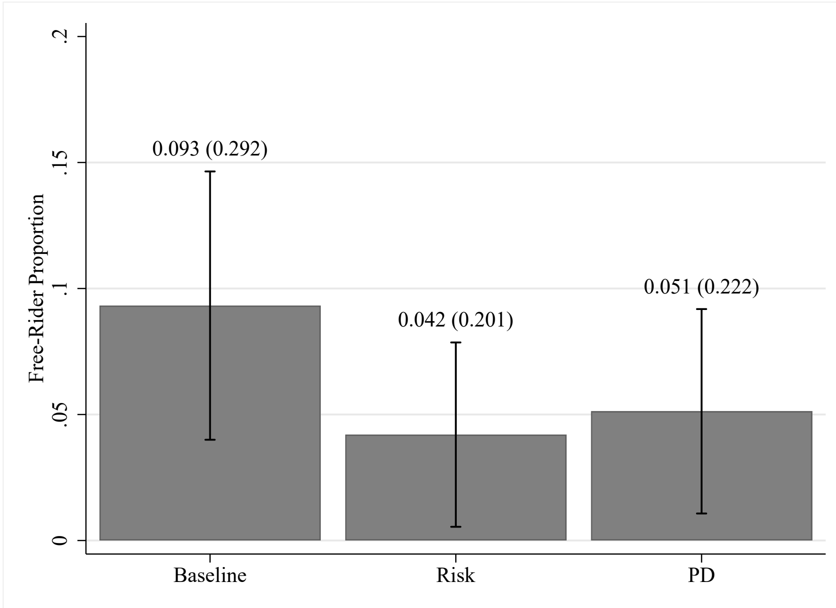
We summarise the findings from the analysis on free-riding behaviour with:

Study 1 - Result 3. *Plausible deniability for zero contributions does not influence the prevalence of free-riding behaviour.*

[14]In table A.2, we repeat the regression analysis only with participants whose fraction of correct answers in the comprehension quiz was above the median within each treatment. The results remain consistent; we find that the *Risk* and *PD* treatments did not impact contributions. Furthermore, in table A.3 we repeat the regression analysis by excluding participants that were non-attentive, did not get all the comprehension questions correct at first try, and who self-reported not to fully understand the instructions as a further robustness check. The results once again remain similar. Note that, excluding participants that did not get the comprehension questions correct at first try is a very strict exclusion criterion in terms of comprehension. This is because participants receive feedback about the correct answers for all comprehension quiz questions and so, conceivably, even if they had initially answered the question incorrectly, they might still have improved their understanding subsequently.

[15]The two robustness check regression analyses that we report in tables A.2 and A.3 in the appendix show similar results.

Figure 3: Proportion of Free-riders – Study 1



Note: We denote the proportion above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

5 Study 2

5.1 Treatments

In study 1, we find that plausible deniability does not influence contributions nor free-riding behaviour. In study 2, we further investigate the robustness of this result. As part of this study, we replicate the *PD* treatment to act as the reference treatment for the two additional treatments: *NoPD* and *PD-Partial*.

The *NoPD* treatment is identical to the *PD* treatment in terms of uncertainty, but does not offer plausible deniability for zero contributions. The parametrisation is kept constant across these two treatments: in both cases, the θ is set to be equal to 0.5 and there is a 20% chance that each player's contribution is lost before reaching the common project. In contrast to the *PD* treatment, in the *NoPD* treatment there is no plausible deniability about contributions of zero: when participants are informed about the contributions of each group member, they are made aware of whether each group members' contribution was lost or not. Figure 4 reproduces an example of how this was presented to the participants. As is clear from the figure, in this instance all group members are aware that player B *intended* to contribute 7 Taler, but this was lost and so the *realised* contribution of player B is 0. Players in the example would also be aware that Player D chose to free-ride, as no contributions are listed as lost on their row. Thus, even though contributions are potentially lost, as in the *PD* treatment, in the *NoPD* treatment participants are not granted plausible deniability when free-riding.

Through comparing *NoPD* to *PD*, we can identify whether the presence of plausible deniability for contributions of zero influences either average contributions or the prevalence of free-riding behaviour. In contrast to the comparisons made in study 1, the comparison between *NoPD* and *PD* in study 2 entails the exact same type of uncertainty, namely uncertainty with regards to whether the intended contributions will fully translate into realised contributions. This allows for

Figure 4: Summary of Contributions on Participant Screens – *NoPD* treatment.

You are player B.

Common Project Contributions	Lost Contributions
Player A: 5 Taler	--
Player B: 0 Taler	7 Taler
Player C: 8 Taler	--
Player D: 0 Taler	--

any differences in behaviour to be clearly attributed to the existence of plausible deniability.

The second new treatment in study 2 is the *PD-Partial* treatment. There is a large literature documenting that dishonest behaviours such as lying tend to manifest only to a limited extent (e.g., Abeler et al., 2019; Fischbacher and Föllmi-Heusi, 2013; Gerlach et al., 2019). This may explain why participants did not exploit the plausible deniability provided in the *PD* treatment in study 1: the reduction in contribution that the average participant would have had to make (approx. 5 Taler) might have been too large of an uncooperative act.

In the *PD-Partial* treatment, instead of providing plausible deniability for contributions of zero, we offer participants plausible deniability for lower contributions, so that they can reduce their contributions by a few units only and still be able to justify that. We implement this through the introduction of a 30% chance that two thirds of each participant's contribution is lost before reaching the common project.[16] As in all of the other treatments, the realisation of contribution loss is independent for each group member. Given these parameters, the expected θ is equal to 0.4, thus, in expectation the public good has the same returns across all the three treatments in study 2. When group members observe the contributions to the common project from each other (as illustrated in figure 1), they cannot be certain if the realised contribution is what was intended, or whether their fellow group members had contributed more but got unlucky and lost two thirds of their intended contribution.

The comparison between the *PD-Partial* and *PD* treatments allows us to identify whether people may exploit plausible deniability only in situations where the extent of the 'immoral' behaviour is limited. We contrast a situation where there is plausible deniability only for zero contributions, in the *PD* treatment, to a situation where plausible deniability offers an excuse for lower contributions than would have otherwise been acceptable, in the *PD-Partial* treatment.

Furthermore, we can compare the *NoPD* treatment with the *PD-Partial* treatment to observe differences in contributions between a situation with no plausible deniability and one with plausible deniability about reductions in contributions (i.e., "partial" contributions) and study how this contrast influences intended contributions and the prevalence of free-riding behaviour. This comparison comes with the caveat that, even though the expected MPCR is constant across the two, the uncertainty between the conversion of intended into realised contributions is of a different flavour. In *NoPD*, one's full intended contribution may be lost, while in *PD-Partial* only two thirds of the intended contribution may be lost. Nevertheless, we believe comparing the two treatments is instructive about any influence of plausible deniability in behaviour in the one-shot public good

[16] For example, a participant who contributed 5 but was unlucky would have had 2 Taler reach the common pot. The calculation for this is, $5 - \frac{2}{3} \times 5 = 1.67$, which is rounded to 2. Following this logic, intended contributions of 10, 9, and 8 would result in a realised contribution of 3; intended contributions of 7, 6, and 5 to a realised contribution of 2; intended contributions of 4, 3, and 2, to a realised contribution of 1; and finally, intended contributions of 1 and 0, to realised contributions of 0.

game.

Implementation Details

The implementation details for study 2 closely match those of study 1: participants were recruited using Prolific, using the same mechanisms to ensure data quality, and the same power analysis to determine the required sample size.[17] The original pre-registration was updated to include the new treatments and the rationale behind them.

Table 2: Participant Demographics and Other Characteristics – Study 2

	PD	*NoPD*	*PD-Partial*
N	125	120	124
Age	39.86	39.31	39.39
Female	0.50	0.49	0.50
Econ Training	0.09	0.18	0.08
Trust in Judicial System	5.48	5.32	5.55
Trust in Government	4.06	4.00	4.37
Trust in Elections	5.35	5.81	5.86
Politically Left(1)-Right(10)	4.64	4.50	4.15
SVO: Competitive	0.01	0.03	0.03
SVO: Individualist	0	0	0
SVO: Prosocial	0.45	0.43	0.50
SVO: Altruist	0.54	0.54	0.47
Risk Aversion	5.82	5.85	5.59

Notes: Econ Training refers to participants reporting that they have trained in economics. The SVO categories report the proportion of participants consistent with each classification. Risk aversion refers to number of safe choices made in the MPL task.

5.2 Results

We report the participant characteristics across the three treatments of study 2 in table 2. There are no significant differences across treatments (all $p > 0.05$, exact MWU test).[18] As in study 1, we find no significant differences across the three treatments in terms of self-reported understanding of the instructions or attentiveness. The collection of the risk preference and pro-sociality measures took place six weeks after the participants played the public good game, and we have 238 out of 369 participants returning. This is roughly an attrition rate of 35%, which is slightly higher than that found in study 1. This is probably due to the longer gap between the studies, but since the main analysis is concerned with the contribution decisions, this is of limited concern.

In figure 5, we report the average contributions across the three treatments of study 2.[19] In the replication *PD* treatment, we find that participants send 5.33 Taler which is remarkably similar

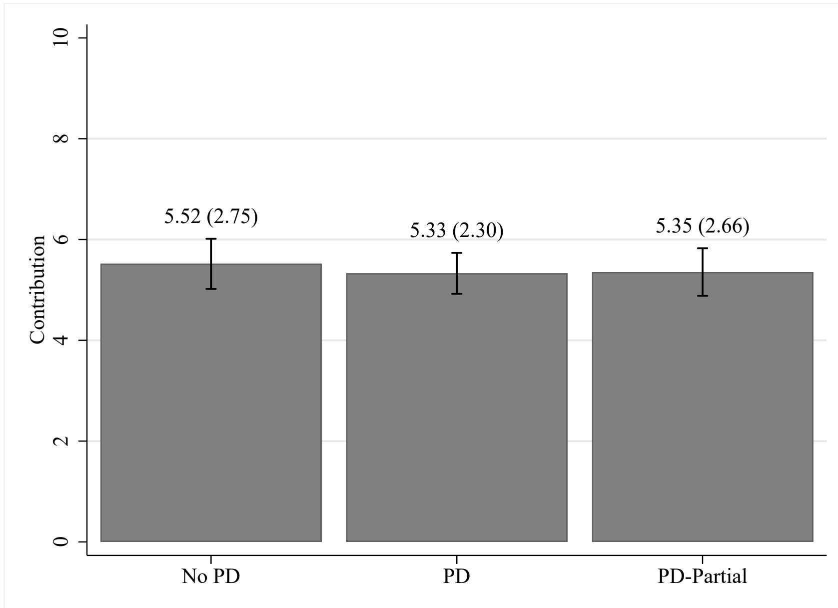
[17] Using the mean and standard deviation of the *PD* treatment in study 1 yields the same required sample size.

[18] The only statistically significant difference at the 5% level is the comparison of participants with economics training between the *PD* and the *PD-Partial* treatments. Given the lack of treatment differences, and in particular between treatments *PD* and *PD-Partial* that have similar proportions of participants with economics training this should not be of concern.

[19] In figure A.4 in the appendix, we depict the full distribution of contributions across treatments.

to the amount sent in the Study 1 *PD* treatment, 5.41 ($p = 0.6970$, exact MWU test). In other words, the behaviour documented in the *PD* treatment in study 1 is replicated in study 2.

Figure 5: Average Contribution – Study 2.



Note: We denote the average *intended* contribution above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

When comparing contributions in *PD* with *NoPD*, we do not observe a statistically significant difference (*PD*: 5.33 vs. *NoPD*: 5.52, $p = 0.5635$, exact MWU test). Hence, it would appear that allowing plausible deniability for contributing nothing does not influence contributions. When looking at contributions in the *PD* and the *PD-Partial* treatment, we do not find a statistically significant difference either (*PD*: 5.33 vs. *PD-Partial*: 5.35, $p = 0.9177$, exact MWU test).[20] Thus, the availability of plausible deniability for lower contributions does not seem to affect the amount that people contribute either. This rules out that the reason why we did not find people were exploiting plausible deniability in study 1, or when comparing the *PD* and *NoPD* treatments here, is because the magnitude of the selfish act would have been too large. We also perform equivalence tests for the comparisons we make in study 2. We find that for both an effect equal to one third of a standard deviation and one half of a standard deviation, the data across the treatments are equivalent.

In table A.4 in the appendix, we report regression analysis where we control for personal characteristics, risk attitudes, pro-sociality as well as the number of correct answers to the comprehension quiz.[21] The regression analysis results are in line with the results of the statistical tests reported above. We do not find a statistically significant difference in contributions between *PD* and *NoPD* or between *PD* and *PD-Partial*.

[20] In the interest of completeness, for the comparison between the *NoPD* and *PD-Partial* treatments we also find no significant treatment differences (*NoPD*: 5.52 vs. *PD-Partial*: 5.35, $p = 0.5830$, exact MWU test).

[21] In table A.5, we repeat the regression analysis only with participants whose fraction of correct answers in the comprehension quiz was above the median within each treatment. The results remain the same. Furthermore, in table A.6 we repeat the regression analysis by excluding participants that were non-attentive, did not get all the comprehension quiz questions correct at first try, and who self-reported not to fully understand the instructions as a robustness check. The results once again remain similar.

We summarise the findings from the analysis of study 2 so far with:

Study 2 - Result 1. *Under an identical implementation of uncertainty, plausible deniability for zero contributions does not induce a change in contributions towards the public good.*

Study 2 - Result 2. *Plausible deniability for lower contributions does not have any effect on contributions towards the public good compared to plausible deniability for zero contributions.*

As argued further above, beyond analysing average contributions to the public good, it is important to also understand if there are any differences in the prevalence of free-riding behaviour across treatments. In the *PD* treatment, participants have plausible deniability for contributions of zero, while in the *NoPD* treatment this is not present. Meanwhile, in the *PD-Partial* treatment, participants have plausible deniability for lower contributions.[22]

In figure 6, we report the proportion of free-riding behaviour across the three treatments. Free-riding behaviour is very similar between *PD* and *NoPD*, with no evidence of any statistical difference ($p = 0.5194$, exact MWU test). Similarly, the proportion of free-riders is statistically indistinguishable between the *PD* and the *PD-Partial* treatments (*PD-Partial*: 0.048 vs. *PD*: 0.040, $p = 1.0000$ exact MWU test). Equivalence tests confirm that for an effect of one half of a standard deviation both of these comparisons are equivalent.[23] In the regression analysis that we report in table A.4 in the appendix, these results are corroborated.[24] We summarise the findings on free-riding behaviour in study 2 with:

Study 2 - Result 3. *Plausible deniability for zero contributions results in no difference in free-riding behaviour.*

Study 2 - Result 4. *Plausible deniability for lower contributions results in no difference in free-riding behaviour.*

6 Discussion & Concluding Remarks

We set out to understand whether the plausible deniability that risk creates may be used as an excuse for individuals to reduce their contributions towards public goods. If this were true, potential inefficiencies in transforming collected proceeds towards a public good (e.g., healthcare or nature conservation) would result in further downward pressure on the willingness of individuals to support and/or contribute towards such schemes. We tackle this question with two online experiments featuring five treatments where we vary whether there is uncertainty in the public good returns, and whether this uncertainty gives participants plausible deniability for full or partial free-riding. We find that neither uncertainty nor the different implementations of plausible deniability influence contributions or the prevalence of free-riding behaviour.

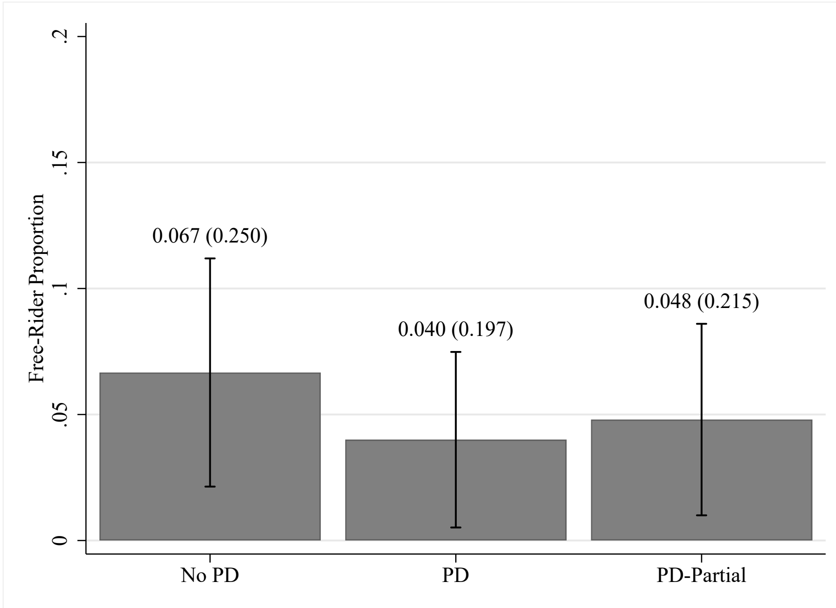
Previous literature studying the impact of uncertainty in the returns of the public good on contributions has yielded mixed theoretical predictions and empirical results. Due to this, we did not formulate a specific directional hypothesis for the effect of uncertainty on contributions.

[22] This also includes realised contributions of zero. For participants in the *PD-Partial* treatment, when observing a contribution of zero from a group member, they could not be sure whether that group member intended to contribute zero or 1.

[23] An equivalence test for an effect of one third of a standard deviation determines the comparison between the *PD-Partial* and *PD* treatments as equivalent, while the comparison between the *PD* and *NoPD* as indeterminate.

[24] Consistently, the two robustness checks we perform, reported in tables A.5 and A.6 in the appendix, find no difference between the treatments.

Figure 6: Proportion of Free-riders – Study 2.



Note: We denote the proportion above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

However, given our hypothesis that people may hide behind the excuses that uncertainty provides, a treatment that would net out the effect of uncertainty from that of plausible deniability was required. Our findings about the effect of uncertainty on contributions are in line with studies that find no impact on average contributions (Levati and Morone, 2013; Brennan et al., 2008; Stoddard et al., 2015; Björk et al., 2016; Théroude and Zylbersztejn, 2020). This supports the risk and other-regarding preference theorisation by Björk et al. (2016). Additionally, it provides reassurance that real life social dilemmas that involve uncertainty may be well captured by standard public good games in the laboratory.

We now turn to our key finding that plausible deniability did not have an effect on cooperation. This result could be observed if there was indeed no difference across our treatments, or there was one that we were unable to detect with our experiment. Through sample size calculations prior to data collection, power calculations, and equivalence tests, we rule out the latter and conclude that if there were undetected differences, these would be too small to be of economic significance.

Since Dana et al. (2007), there have been a plethora of replications of the impact of moral wiggle room on allocation decisions. To the best of our knowledge, the present study is the first to investigate how plausible deniability manipulations might impact behaviour in the public good game. It may be that the effect, which has been widely replicated in non-strategic settings such as modified dictator games, does not consistently extend to strategic contexts such as the social dilemmas captured by the public good game. This is congruent with the mixed findings from tests of moral wiggle room in the trust game. Van der Weele et al. (2014) investigate how plausible deniability might influence second mover behaviour in the trust game and find no effect. They discuss that the reciprocal nature of the trust game offers more morally relevant information about interacting partners, while also triggers motivations that are less malleable to plausible deniability. Our one-shot implementation of the public good game rules out any effect of reciprocity by design, thus, the lack of an effect in our study suggests that contexts or situations that do not entail reciprocity might also be immune to moral wiggle room manipulations. In contrast to Van der Weele et al. (2014), Regner and Matthey (2021) and Regner and Matthey (2021) do find evidence of moral wiggle room effects in the trust game.

The number of people involved in the decision could be a factor affecting whether people decide to hide behind excuses to protect their image. Whilst the exploitation of moral wiggle room has been widely observed in two-person interactions, where the level of exposure and scrutiny is high, our study provides evidence that this may not extend to multi-people contexts. It could be that as the number of people increases, the scope for social image to influence decisions is diluted. This effect would be akin to the diffusion of responsibility phenomenon in the bystander effect (Darley and Latané, 1968; Latané and Nida, 1981). It could also be obtained more mechanically: as the number of people involved increases, the importance of how each of them sees you is diluted (Latané, 1981), thus, reducing the influence of plausible deniability in a multi-person context. Exploring further the interplay between the number of interacting players and the influence of moral wiggle room would be an interesting extension of the present study.

An additional possible explanation for the absence of an effect of plausible deniability could be related to the higher complexity, and therefore increased attention that is required to make decisions in strategic settings. Consistent with the intuition that finite attention limits the weight that each attribute can play in multi-attribute choice (e.g., Sims, 2003), it could be that considering strategic uncertainty, natural uncertainty, and image concerns is too demanding for participants, limiting the role that the latter, or perhaps the last two, could have on decisions in strategic settings like the public good game.

Whilst it is a possibility, it is unlikely that our design is driving our results. We have reasons to believe that the null effect is not driven by a lack of participant awareness or understanding about our plausible deniability manipulation. The public good game can be quite a complex game, especially for those interacting with it for the first time. Despite this, a large proportion of our participants evidenced their understanding through correct comprehension questions (the range of average correct answers is $74\% - 84\%$ across all treatments in both studies).[25] It is unlikely that this understanding applied to the game instructions but excluded the manipulation. Indeed, we perform robustness analysis to ensure our null effect is not due to any issues of comprehension or attention. We re-estimate our regressions either with only participants that were above median in terms of correct answers in the comprehension quiz, or by excluding any participants that were identified as non-attentive or self-reported not to fully understand the instructions. In both cases the results remain consistent, indicating that lack of understanding or attention are unlikely to be driving our null results.[26]

Furthermore, 73.5% of participants across both studies answered correctly a question about the number that would appear on the contributions table if they were unlucky and their contribution was lost. This shows that they were aware that a 0 would appear not only if people did not contribute, as in standard public good games, but also if they are unlucky.[27] Note that all participants received feedback with the correct answer, so even those that did not respond correctly initially may have learned about it afterwards. The feedback itself may have increased the salience of the manipulation, thus, increasing awareness of it amongst participants. Asking participants any more specific manipulation checks could have jeopardised its potential effect, as such questions would

[25] In fact, in all treatments but two, the median score in the comprehension quiz was 100%. In the *Risk* treatment in study 1 this was 80% and in the *PD* treatment in study 2 this was 75%.

[26] See tables A.1 - A.3 in the Appendix for the main regression and two robustness analyses for study 1 and tables A.4 - A.6 for the equivalent tables for study 2.

[27] In table A.7 in the appendix, we perform a further robustness check where we re-run our regressions with only those participants that answered this question correctly and once again find similar null results.

essentially force participants to admit that they are aware there is plausible deniability to exploit, leaving their social image no longer protected. Instead of a direct question to check awareness of the manipulation, we include a sentence in the instructions to make the manipulation explicit: *"if you see someone's contribution is 0, you will not be able to tell whether they invested 0 or whether their contribution got lost"*. Thus, we believe that lack of understanding, or unawareness of the available plausible deniability to be exploited are unlikely to have driven the null results that we report.

Finally, another important difference between our and previous work investigating the impact of plausible deniability concerns implementation: whilst previous studies have largely been implemented in physical laboratories, our study was conducted online. This is because online studies have the advantage of reaching a much more representative sample of the population, which allows us to obtain more generalisable results. We do not believe, however, that this could be behind the difference in results.

In a recent study closely resembling ours in terms of implementation (online labour marketplace Prolific and one-shot PGGs), Struwe et al. (2024) report very similar decisions online and in the lab. Arechar et al. (2018) report similar overall patterns of behaviour online and in the lab as well but highlight that, without punishment, cooperation was higher online. Higher baseline levels of cooperation should give our hypothesised effect a wider scope to materialise online than it would in a physical lab, with the absence of an effect in our study hence suggesting we would not find it in the lab either. If we were to find the effect in the lab, it would come with serious questions regarding its generalisability.

Further suggesting that this methodological difference is unlikely to be driving our results are the following studies, which were either conducted online and found an effect, or in the lab and did not find one. First, Exley (2020)'s study 2 is an online experiment and similar evidence of excuse-driven impact on charitable giving is reported. Second, Van der Weele et al. (2014) was implemented in the laboratory and did not find any effect coming from moral wiggle room manipulations.

Overall, the finding that inefficiencies in-built in the provision of public goods do not impact contributions is a prospectively hopeful message. Given the overwhelming evidence on how challenging it is to sustain cooperation, identifying that government inefficiencies need not further jeopardise the provision of public goods and services is reassuring. Of course, this conclusion stems from the specific context that we study here, and future research should explore this further either with alternative manipulations or different subject pools. For example, understanding whether participants from areas where government inefficiencies or corruption are typically more blatant would be more likely to succumb to any plausible deniability would be an interesting extension of the present study. We hope that our results inspire further studies in this direction.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

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

BBC (2022). *Government inefficiency costs £35bn a year, says ex-minister.* `https://www.bbc.co.uk/news/uk-politics-63131393` [Accessed: 27/11/2024].

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.

Bigoni, M. and Suetens, S. (2012). Feedback and dynamics in public good experiments. *Journal of Economic Behavior & Organization*, 82(1):86–95.

Björk, L., Kocher, M., Martinsson, P., and Nam Khanh, P. (2016). Cooperation under risk and ambiguity. *Working Paper Series, University of Gothenburg, No 683.*

Bodner, R. and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The psychology of economic decisions*, 1(105):26.

Brennan, G., González, L. G., Güth, W., and Levati, M. V. (2008). Attitudes toward private and collective risk in individual and strategic choice situations. *Journal of Economic Behavior & Organization*, 67(1):253–262.

Brewer, M. B. and Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of personality and social psychology*, 50(3):543.

Brock, J. M., Lange, A., and Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, 103(1):415–437.

Burton-Chellew, M. N. and West, S. A. (2021). Payoff-based learning best explains the rate of decline in cooperation across 237 public-goods games. *Nature human behaviour*, 5(10):1330–1338.

Cagala, T., Glogowsky, U., Grimm, V., and Rincke, J. (2019). Public goods provision with rent-extracting administrators. *The Economic Journal*, 129(620):1593–1617.

Cárdenas, J.-C., Janssen, M. A., Ale, M., Bastakoti, R., Bernal, A., Chalermphol, J., Gong, Y., Shin, H., Shivakoti, G., Wang, Y., et al. (2017). Fragility of the provision of local public goods to private and collective risks. *Proceedings of the National Academy of Sciences*, 114(5):921–925.

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

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

Christens, S., Dannenberg, A., and Sachs, F. (2019). Identification of individuals and groups in a public goods experiment. *Journal of Behavioral and Experimental Economics*, 82:101445.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33:67–80.

Dannenberg, A., Johansson-Stenman, O., and Wetzel, H. (2022). Status for the good guys: An experiment on charitable giving. *Economic Inquiry*, 60(2):721–740.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., and Tavoni, A. (2015). On the provision of public goods with probabilistic and ambiguous thresholds. *Environmental and Resource economics*, 61:365–383.

Dannenberg, A., Lumkowsky, M., Carlton, E. K., and Victor, D. G. (2023). Naming and shaming as a strategy for enforcing the paris agreement: The role of political institutions and public concern. *Proceedings of the National Academy of Sciences*, 120(40):e2305075120.

Darley, J. M. and Latané, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology*, 8(4p1):377.

Dickinson, D. L. (1998). The voluntary contributions mechanism with uncertain group payoffs. *Journal of economic behavior & organization*, 35(4):517–533.

Drouvelis, M., Lohse, J., and Qiu, Z. (2025). Transparency and whistleblowing in public goods provision with rent-extracting aadministrators. *Available at SSRN 5586674*.

Dufwenberg, M., Gneezy, U., Güth, W., and van Damme, E. (2001). Direct versus indirect reciprocity: An experiment. *Homo Oeconomicus-Journal of Behavioral and Institutional Economics*, 18(1/2):19–30.

Ebeling, F. (2013). On the role of endowment heterogeneity and ambiguity for conditional cooperation. *Working Paper Series: Cologne University No. 58*.

Ellingsen, T. and Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *American economic review*, 98(3):990–1008.

Exley, C. L. (2016). Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*, 83(2):587–628.

Exley, C. L. (2020). Using charity performance metrics as an excuse not to give. *Management Science*, 66(2):553–563.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. *Journal of Economic Psychology*, 45:253–267.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American economic review*, 100(1):541–556.

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

Fischbacher, U., Schudy, S., and Teyssier, S. (2014). Heterogeneous reactions to heterogeneity in returns from public goods. *Social Choice and Welfare*, 43:195–217.

Flambard, V., Le Lec, F., and Romaniuc, R. (2024). An experimental comparison of contributions in collective prevention games and public goods games. *Economic Inquiry*, 62(4):1598–1617.

Freundt, J. and Lange, A. (2021). On the voluntary provision of public goods under risk. *Journal of Behavioral and Experimental Economics*, 93:101727.

Gangadharan, L. and Nemes, V. (2009). Experimental analysis of risk and uncertainty in provisioning private and public goods. *Economic Inquiry*, 47(1):146–164.

Garcia, T., Massoni, S., and Villeval, M. C. (2020). Ambiguity and excuse-driven behavior in charitable giving. *European Economic Review*, 124:103412.

Gerlach, P., Teodorescu, K., and Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin*, 145(1):1.

Grandjean, G., Lefebvre, M., and Mantovani, M. (2022). Preferences and strategic behavior in public goods games. *Journal of Economic Behavior & Organization*, 195:171–185.

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

Grossman, Z. and Van Der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.

Guth, W., Huck, S., and Ockenfels, P. (1996). Two-level ultimatum bargaining with incomplete information: An experimental study. *The Economic Journal*, 106(436):593–604.

Herrmann, B., Thoni, C., and Gachter, S. (2008). Antisocial punishment across societies. *Science*, 319(5868):1362–1367.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

Huang, J., Kessler, J. B., and Niederle, M. (2024). Fairness has less impact when agents are less informed. *Experimental Economics*, 27(1):155–174.

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social psychological and personality science*, 8(4):355–362.

Lambrecht, M., Proto, E., Rustichini, A., and Sofianos, A. (2024). Intelligence disclosure and cooperation in repeated interactions. *American Economic Journal: Microeconomics*, 16(3):199–231.

Latané, B. (1981). The psychology of social impact. *American psychologist*, 36(4):343.

Latané, B. and Nida, S. (1981). Ten years of research on group size and helping. *Psychological bulletin*, 89(2):308.

Levati, M. V. and Morone, A. (2013). Voluntary contributions with risky and uncertain marginal returns: the importance of the parameter values. *Journal of Public Economic Theory*, 15(5):736–744.

Levati, M. V., Morone, A., and Fiore, A. (2009). Voluntary contributions with imperfect information: An experimental study. *Public Choice*, 138:199–216.

Matthey, A. and Regner, T. (2011). Do i really want to know? a cognitive dissonance-based explanation of other-regarding behavior. *Games*, 2(1):114–135.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. J. (2011). Measuring social value orientation. *Judgment and Decision making*, 6(8):771–781.

NHS Counter Fraud Authority (2024). What is the cost of nhs fraud? `https://cfa.nhs.uk/` [Accessed: 27/11/2024].

Nisan, M. and Horenczyk, G. (1990). Moral balance: The effect of prior behaviour on decision in moral conflict. *British journal of social psychology*, 29(1):29–42.

Oechssler, J., Reischmann, A., and Sofianos, A. (2022). The conditional contribution mechanism for repeated public goods–the general case. *Journal of Economic Theory*, 203:105488.

Ploner, M. and Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93:374–383.

Regner, T. (2018). Reciprocity under moral wiggle room: Is it a preference or a constraint? *Experimental Economics*, 21(4):779–792.

Regner, T. and Matthey, A. (2021). Actions and the self: I give, therefore i am? *Frontiers in psychology*, 12:684078.

Reischmann, A. and Oechssler, J. (2018). The binary conditional contribution mechanism for public good provision in dynamic settings—theory and experimental evidence. *Journal of Public Economics*, 159:104–115.

Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of pharmacokinetics and biopharmaceutics*, 15:657–680.

Seçilmiş, E. (2018). An experimental analysis of moral self-regulation. *Applied economics letters*, 25(12):857–861.

Serra-Garcia, M. and Szech, N. (2022). The (in) elasticity of moral ignorance. *Management Science*, 68(7):4815–4834.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Stoddard, B. V. et al. (2015). Probabilistic production of a public good. *Economics Bulletin*, 35(1):37–52.

Struwe, N., Blanco, E., and Walker, J. M. (2024). Increasing benefits in one-time public goods does not promote cooperation. *Proceedings of the National Academy of Sciences*, 121(41):e2410326121.

Théroude, V. and Zylbersztejn, A. (2020). Cooperation in a risky world. *Journal of Public Economic Theory*, 22(2):388–407.

Thöni, C. and Volk, S. (2018). Conditional cooperation: Review and refinement. *Economics Letters*, 171:37–40.

Van der Weele, J. J., Kulisa, J., Kosfeld, M., and Friebel, G. (2014). Resisting moral wiggle room: how robust is reciprocal behavior? *American economic Journal: microeconomics*, 6(3):256–264.

Van Dijk, E., De Cremer, D., and Handgraaf, M. J. (2004). Social value orientations and the strategic use of fairness in ultimatum bargaining. *Journal of experimental social psychology*, 40(6):697–707.

The Limits of Moral Wiggle Room:Plausible Deniability in Public Good Games

Online Appendix

Danae Arroyos-Calvera & Andis Sofianos

May 19, 2026

A Experimental Instructions

Contents:

- Instructions
- Part 2: MPL & SVO
- Comprehension Questions by treatment

Key:

- Baselines / "T1" refers to the *Risk* treatment / "T2", "T2x" refer to *PD* (*Plausible Deniability*) and *NoPD* respectively; and "T3", to *PD-Partial*.
- T1 order 1 is where "theta = 0.3" is presented before "theta = 0.5", and the example has theta = 0.3. Order 2 is the opposite.
- **[Comments to reader]**
- "______________" indicate page break
- Page titles. When between "[]", participants did not see a title.

Instructions

Informed Consent

[5 min timer]

Welcome to this study!

Upon completing the online study, you will receive a £1 participation fee. You can earn additional money based on the decisions you make during the study and luck. It is therefore very important that you read all instructions with care.

There are no risks associated with this study beyond those associated with normal computer use. The benefits will include the monetary award that you will receive shortly after completing the study.

All information you give to the investigator is confidential. The records of this study will be kept private. No identifiers linking you to the study will be included in any sort of report that might be published. Research records will be stored securely and only [redacted] and [redacted] will have access to the records. If you wish to withdraw your data from the study, please email the researchers on the email address below giving your Prolific ID and stating your wish to withdraw your data. This will be possible up to 2 weeks from the date of the first session (i.e. for the duration of the data collection plus one more week).

Your decision to participate will not affect your current or future relations with the Universities of [redacted] or [redacted]. If you decide to participate, you are free to withdraw from the study at any time. There will be no adverse consequences for you if you withdraw from the study, other than forfeiting your payment.

You can contact the experimenters with any questions or comments regarding the study using the Prolific messaging system.

You understand that this research study has been reviewed by the [redacted] and the Humanities & Social Sciences Ethical Review Committee, University of [redacted].

Consent

- ☐ I have read and understood the information provided
- ☐ I agree to voluntarily participate in this study

____

[ProlificID]

[3 min timer]

Write (or copy-paste) your Prolific ID here:

____

Information

[4 min timer]

Upon completing this study, you will receive an initial payment of £1. Your additional bonus depends on decisions in the tasks you will be asked to complete. Details about this will be offered later on.

In order to avoid excessive waiting time, all participants will have a maximum of 10 minutes per screen. This is enough time, you will not have to rush.

If you take more than 10 minutes in any screen, you will be automatically kicked out of the study and you will not be able to continue or re-start it. If that happens, please return your submission **[last wave, due to Prolific changes]** by clicking on “Stop Without Completing”.

If you do not have at least 15 minutes to complete this study right now, please abandon it by closing your browser now and return your submission.

____

Welcome to our Study

[10 min timer]

After you complete the study, you will receive £1.00 as a participation fee plus your bonus earnings from this experiment.

- **[Baselines]** Your earnings will depend on your own decisions and the decisions of other participants. It is therefore important that you read these instructions carefully.
- **[T1, T2, T2x, T3]** Your earnings will depend on your own decisions, the decisions of other participants and chance. It is therefore important that you read these instructions carefully.
- The currency for the experiment will be Taler. At the end, each Taler will be exchanged for £0.10.

Instructions

[10 min timer]

In this part of the study, you will be part of a group of four participants.

Each participant starts with 10 Taler in their private account. You (as well as all other group members) need to decide how many of those 10 Taler you want to invest into a common project. Everyone will make this decision at the same time.

You and the other group members will decide anonymously. That is, other group members will see each group member's contribution to the common project, but they will not be able to match them with your identity or that of any group member.

The common project will generate income for everyone in your group. For each Taler contributed to the common project, each member of the group will receive **[baseline_0.4]** 0.4 **/ [baseline_0.5]** 0.5 **/ [T1: order1]** either 0.3 or 0.5 **/ [T1: order2]** either 0.5 or 0.3 **/ [T2, T2x, T3]** 0.5 Taler. **[T1]** Whether you and the other members of your group receive **[T1: order1]** 0.3 or 0.5 **/ [T1: order2]** 0.5 or 0.3 Taler for each Taler contributed into the common project will depend on chance. It is equally likely that your group will receive **[T1: order1]** either 0.3 or 0.5 **/ [T1: order2]** either 0.5 or 0.3.

You can think of this as a lottery game where there are 100 balls in an opaque bag: 50 of them (50%) are blue and 50 of them (80%) are purple. One ball is drawn without looking. If the ball is blue, you and the other members of your group will receive **[T1: order1]** 0.3 Taler for each Taler contributed into the common project; if the ball is purple, the return will be 0.5. **[T1: order2]** 0.5 Taler for each Taler contributed into the common project; if the ball is green, the return will be 0.3.

You will receive the Taler that you choose not to invest into the common project (i.e., the Taler that are left in your private account).

In summary,
[T1] If your group receives **[T1: order1]** 0.3 **/ [T1: order2]** 0.5 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be ***[T1: order1]*** *0.3* ***/ [B_0.4]*** *0.4* ***/ [B_0.5, T1: order2 and T2, T2x, T3]*** *0.5 x (total group contribution to the common project)*

- Each participant's total earnings will be
  *(10 – own investment) + **[T1: order1]** 0.3 **/ [B_0.4]** 0.4 **/ [B_0.5, T1: order2 and T2, T2x, T3]** 0.5 x (total group contribution to the common project)*

[T1] If your group receives **[T1: order1]** 0.5 **/ [T1: order2]** 0.3 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be
  *(**[T1: order1]** 0.5 **/ [T1: order2]** 0.3 x (total group contribution to the common project)*
- Each participant's total earnings will be
  *(10 – own investment) + **[T1: order1]** 0.5 **/ [T1: order2]** 0.3 x (total group contribution to the common project)*

Comprehension Question

What is the chance that you and the other members of your group receive 0.3 Taler for every Taler invested into the common project?

- No chance
- 20% chance
- ✓ 50% chance
- 80% chance

[This is an attention check.]

[SomeExamples]

[10 min timer]

Read carefully the following examples as there will be some comprehension questions right after this screen.

[The following box is only shown if they click on 'Show payment info reminder' button]

Remember that,

[T1] If your group receives **[T1: order1]** 0.3 **/ [T1: order2]** 0.5 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be
  ***[T1: order1]** 0.3 **/ [B_0.4]** 0.4 **/ [B_0.5, T1: order2 and T2, T2x, T3]** 0.5 x (total group contribution to the common project)*
- Each participant's total earnings will be
  *(10 – own investment) + **[T1: order1]** 0.3 **/ [B_0.4]** 0.4 **/ [B_0.5, T1: order2 and T2, T2x, T3]** 0.5 x (total group contribution to the common project)*

[T1] If your group receives **[T1: order1]** 0.5 **/ [T1: order2]** 0.3 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be [(**[T1: order1]** *0.5* **/ [T1: order2]** *0.3 x (total group contribution to the common project)]*
- *Each participant's total earnings will be (10 – own investment) +* **[T1: order1]** *0.5* **/ [T1: order2]** *0.3 x (total group contribution to the common project)*

[T1: order1]

If your group receives 0.3 Taler for each Taler contributed to the common project:

Example 1: You invest 4 Taler into the project, 2 other participants each invested 8 Taler, and 1 other participant invested 0 Taler. Thus, a total of *4 + 8 + 8 =* 20 Taler are invested in the common project. You will receive a total of *20 x 0.3 =* 6 Taler from the common project. You will also receive the 6 Taler you kept in your private account. So you earn a total of 12 Taler.

Example 2: All participants in the group, including yourself, invest their 10 Taler into the common project. Thus, a total of 40 Taler are invested in the common project. You will receive a total of *40 x 0.3 =* 12 Taler from the common project. You will not receive any Taler from your private account since you invested them all. So you earn a total of 12 Taler.

Example 3: All participants in the group, including yourself, invest 0 Taler into the common project. Thus, a total of 0 Taler are invested in the common project. You will receive nothing from the common project. You will receive the 10 Taler you kept in your private account. So you earn a total of 10 Taler.

Example 4: All the other participants invest 10 Taler each into the common project, and you invest nothing. Thus, a total of 30 Taler are invested in the common project. You will receive a total of *30 x 0.3 =* 9 Taler from the common project. You will also receive the 10 Taler you kept in your private account. So you earn a total of 19 Taler.

Example 5: All the other participants invest nothing into the common project, and you invest 10 Taler. Thus, a total of 10 Taler are invested in the common project. You will receive a total of *10 x 0.3* = 3 Taler from the common project. You will not receive any Taler from your private account since you invested them all in the common project. So you earn a total of 3 Taler.

If your group is rewarded with 0.5 Taler for each Taler contributed to the common project:

[The same 5 examples were shown here, but with theta = 0.5 instead of 0.3. The change in theta will also change the numbers that depend on it.]

[T1: order2]

If your group is rewarded with 0.5 Taler for each Taler contributed to the common project:

[The same set of 5 examples were shown here, but with theta = 0.5. The change in theta will also change the numbers that depend on it.]

If your group is rewarded with 0.3 Taler for each Taler contributed to the common project:

[The same 5 examples were shown here, but with theta = 0.3. The change in theta will also change the numbers that depend on it.]

[NB: In the other conditions, as there is only one possible theta, participants will only see one set of examples. The change in theta will also change the numbers that depend on it.]

____

Comprehension Questions

[10 min timer]

[T1] Imagine / **[B_0.4, B_0.5, *T2, T2x, T3*]** Remember that the common project gives each group member **[B_0.4]** 0.4 / **[T1: order2]** 0.3 / **[B_0.5, T1: order1 and T2]** 0.5 Taler for every Taler in the common project.

[NB: This is the reverse order of appearance than later on, for variety. The order of the answers is randomised.]

How many Taler would each group member receive as income from the common project?

- ✓ [Theta x Total group contribution to the common project] Taler
- [Theta x Own contribution to the common project] Taler
- [Theta x Others' contribution to the common project] Taler

[NB: Theta will be 0.5 if order1 and 0.3 if order2 in t1, and 0.5 in t2.]

Your total earnings would then amount to **[same as above]**

- ✓ [(Theta x Total group contribution to the common project) + Taler in private account] Taler
- [Own contribution + (Theta x Others' contribution to the common project)] Taler
- [Theta x (Total group contribution to the common project)] Taler
- [Own contribution + (Theta x Others' contribution) + Taler in private account] Taler

____

[Feedback_words]

[10 min timer]

[If both answers are right] That's right!

[If not] **That isn't completely right.**

Remember that,

[T1] There is equal chance that the return is **[T1: order 1]** 0.5 and 0.3 / **[T1: order 2]** 0.3 and 0.5 Taler for each Taler in the common project, but imagine that for these questions chance has determined that the return is **[T1: order 1]** 0.5 / **[T1: order 2]** 0.3.

- Each group member would receive "**[B_0.4]** *0.4* / **[T1: order 2]** *0.3* **/ [B_0.5, T1: order1 and T2]** *0.5 x (total group contribution)*" Taler as income from the common project.
- And your total earnings would amount to "*[***[B_0.4]** *0.4* / **[T1: order 2]** *0.3* **/ [B_0.5, T1: order1 and T2]** *0.5 x (total group contribution)] + Taler in private account (not invested)*" Taler.

____

Comprehension Questions

[10 min timer]

Imagine you invest 7 Taler into the common project, and the total contribution to the common project from the other three group members is 13 Taler.

[T1] Imagine **/ [B_0.4, B_0.5, *T2, T2x, T3*]** Remember that the common project gives each group member **[B_0.4]** 0.4 / **[T1: order 1]** 0.3 **/ [B_0.5, T1: order2 and *T2, T2x, T3*]** 0.5 Taler for every Taler in the common project.

[T1] Remember that there is equal chance that the return is / **[T1: order 1]** 0.3 and 0.5 / **[T2: order 1]** 0.5 and 0.3, but for these questions chance has determined that it is **[T1: order 1]** 0.3 / **[T2: order 1]** 0.5.

How many Taler would each group member receive as income from the common project? **[The actual options are numbers rather than text, computed using the following expressions]**

- ✓ **[Theta x (Total group contribution to the common project)]** Taler **[quiz3_ans1]**
- **[Theta x Own contribution]** Taler
- **[Theta x Others' contribution]** Taler

Your total earnings would then amount to **[same as above]**

- ✓ **[Theta x (Total group contribution to the common project) + Taler in private account]** Taler **[quiz4_ans1]**
- **[Own contribution + (Theta x Others' contribution)]** Taler
- **[Theta x (Own + Others' contribution)]** Taler
- **[Own contribution + (Theta x Others' contribution) + Taler in private account]** Taler

Choose five taler in this question… **[attention check – participants who fail it cannot proceed; in T2x and T3, Prolific regulations with regards to attention checks had changed, and participants could only be rejected if they had failed two attention checks].**

- 12.0 Taler
- 7.5 Taler
- ✓ 5.0 Taler
- 0.0 Taler

Failed Attention Check

[only displayed if they fail the attention check]

A question in the previous screen was an attention test and you did not pass it.

It is very important for the quality of our data that you are paying enough attention.
Please make sure to pay attention and read questions carefully in the rest of the study.
Thank you for your cooperation.

[Feedback_new]

[10 min timer]

[If both answers are right] That's right!

[If not] **That isn't completely right.**

[T1] Remember that there is equal chance that the return is / **[T1: order 1]** 0.3 and 0.5 / **[T1: order 2]** 0.5 and 0.3.

Remember that,

- There is equal chance that the return is **[T1: order 1]** 0.3 and 0.5 / **[T1: order 2]** 0.5 and 0.3, but for these questions chance determined that it was is **[T1: order 1]** 0.3. / **[T1: order 2]** 0.5.
- Each group member would receive "**[B_0.4]** *0.4* / **[T1: order 1]** *0.3* **/ [B_0.5, T1: order2 and** ***T2, T2x, T3*****]** *0.5 x (total group contribution)*" as income from the common project.
- And your total earnings would amount to "*[***[B_0.4]** *0.4* / **[T1: order 1]** *0.3* **/ [B_0.5, T1: order2 and** ***T2, T2x, T3*****]** *0.5 x (total group contribution)] + Taler in private account (not invested)*".

The returns of the common project would be **[B_0.4]** 0.4 / **[T1: order 1]** 0.3 **/ [B_0.5, T1: order2 and** ***T2, T2x, T3*****]** 0.5 Taler for every Taler in the common project.

- Each group member would receive *(7 + 13) x* ***[group.theta]*** = **[player.quiz3_ans1]** Taler from the common project.

- Your total income would then amount to **[*player.quiz3_ans1]* + 3 = [player.quiz4_ans1]** Taler.

This is how the group contributions to the common project will be shown to you and your group members. Imagine you are player B.

You are player B.

Common Project Contributions
Player A: 5 Taler
Player B: 7 Taler
Player C: 8 Taler
Player D: 0 Taler

[Treatments T2, T2x, T3 only]
[10 min timer]

Investments may get lost

There is a [T2, T2x] 20% [T3] 30% chance that [T3] a part (two thirds) of the investment to the common project of each of your group members (including yourself) is lost. You can think of this as a lottery game where there are 100 coloured balls in an opaque bag: [T2, T2x] 20 of them (20%) [T3] 30 of them (30%) are red and [T2, T2x] 80 of them (80%) [T3] 70 of them (70%) are green. One ball is drawn without looking. If the ball is red, [T3] two thirds of the investment is lost, and if the ball is green, [T3] none of the investment is lost [T2,T2x] the investment is not lost.

This [T2, T2x] 20% [T3] 30% chance is independent for each group member. You can imagine this as each group member (including yourself) having their own opaque bag of balls to determine whether [T3] two thirds of their investment is lost.

[T2x] Every group member (including yourself) will find out whether this happened to their own investment and everyone else's. [T2, T3] Every group member (including yourself) will find out whether this happened to their own investment after making it. Nobody will find out whether this happened to the investment of any other group member. [T2] This means that if you see someone's contribution is 0, you will not be able to tell whether they invested 0 or whether their contribution got lost. [T3] This means that when you see someone's contribution, you will not be able to tell whether that is the amount they invested or whether part of their contribution got lost.

Recall the previous example:

Imagine you invest 7 Taler into the common project, and the total contribution to the common project from the other three group members is 13 Taler.

Remember that the common project gives each group member 0.5 Taler for every Taler in the common project.

Imagine that [T3] two thirds of your investment got lost. The previous example would then be as follows.

Imagine you invest 7 Taler into the common project **but [T3] two thirds of your investment gets lost and never reaches the common project**. **[T3] This means that out of your 7 Taler contribution, only 2 (rounded from 2.33) reach the common pot.** Imagine that the total contributions to the common project from the other three group members is 13 Taler. **[T3] This would make a total of 15 Taler in the common pot.**

Remember that the common project gives each group member 0.5 Taler for every Taler in the common project.

This is how the group contributions to the common project will be shown to you and your group members. Imagine you are player B.

You are player B.	**[this column shown in T2x only]**
Common Project Contributions	**Lost Contributions**
Player A: 5 Taler	--
Player B: 0 Taler **[T3] 2 Taler**	7 Taler
Player C: 8 Taler	--
Player D: 0 Taler	--

You would receive from the common project a total of **[T2,T2x]** *13 x 0.5* = 6.5 **[T3]** (13+2) x 0.5 = 7.5 Taler. You would also keep the 3 Taler you did not invest in the common project. So you would earn a total of **[T2, T2x]** *6.5* **[T3]** *7.5 + 3* = **[T2, T2x]** 9.5 **[T3]** 10.5 Taler.

Comprehension Questions

[quiz_chance_other] What is the chance that [T3] two thirds of the investment of each of your group members is not lost?

- No chance
- 20% chance
- 30% chance
- 50% chance
- ✓ 70% chance [correct answer in T3]
- ✓ 80% chance [correct answer in T2 and T2x]

[quiz_chance_you] What is the chance that [T3] two thirds of your investment is lost? **[attention check]**

- No chance
- ✓ 20% chance [correct answer in T2, T2x]
- ✓ 30% chance [correct answer in T3]
- 50% chance
- 70% chance
- 80% chance

[quiz_investment] If you invested 7 Taler but **[T3]** two thirds of it **[T2, T2x]** they got lost, what contribution would appear from you as Player B on the Common Project Contributions table?

- ✓ 0 Taler [correct answer in T2, T2x]
- ✓ 2 Taler [correct answer in T3]
- 5 Taler
- 7 Taler
- It is impossible to tell from the information provided
- A different amount

Failed Attention Check

[same 'failed attention check' message as before is displayed if this is the first attention check that they fail; if this is their second fail, then they get the following message and there is no NEXT button on the screen]

A question in the previous screen was an attention test and you did not pass it.

This was the second attention check you failed. Because of that, you cannot continue with the study.

Please, close your browser and return your submission **[last wave, due to Prolific changes]** by clicking on "Stop Without Completing".

[Feedback_two]

[10 min timer]

[If both answers to the comprehension questions are right] That's right!

[If not] **That isn't completely right.**

Remember that there is a **[T2, T2x]** 20% **[T3]** 30% chance that **[T3]** a part (two thirds) of the investment to the common project of each of your group members (including yourself) is lost. This is as if each of you had an opaque ball with 100 coloured balls, **[T2, T2x]** 20 of which (20%) **[T3]** 30 of which (30%) were red losing balls, and **[T2, T2x]** 80 of which (80%) **[T3]** 70 of which (70%) were green non-losing balls.

For each of your group members, there is a **[T2, T2x]** 20% **[T3]** 30% chance of drawing a red ball and losing their investment, and therefore **[T2, T2x]** an 80% **[T3]** a 70% chance of drawing a green ball and not losing **[T3]** any of their investment.

The same is true for your own chances: there is a **[T2, T2x]** 20% **[T3]** 30% chance of drawing a red ball and losing your investment, and therefore an **[T2, T2x]** 80% **[T3]** 70% chance of drawing a green ball and not losing **[T3]** any of your investment.

[T2, T2x] If your investment or that of one of your group members is lost, it will appear on the contributions table as a contribution of 0, as 0 Taler reach the common pot. **[T3]** If two thirds of your investment were lost, it will appear on the contributions table as a contribution of 2, as 7 minus two thirds of 7 is equal to 2 (rounded from 2.3), so 2 Taler reach the common pot.

[Attention check]

[10 min timer]

Investment Decision

This is a screen to ensure you are attentive to the instructions and the questions asked during this task. Please answer the question below with the result of “four minus one”.

There are 10 Taler in your private account. What is the answer to the question above?

[dropdown list 0-10]

[Attention check feedback 2]

Attention check

[10 min timer]

The previous question was an attention check and you

[failed attention check] did not pass it.

That may mean that you were not paying enough attention while reading the instructions. Please, make sure to be fully focused on this study from now on.

The next screen will ask you about your investment decision. If you want, you can instead go back and go through the instructions and comprehension questions again. If so, select that option below and click on the ‘Next’ button.

☐ Take me back to see the instructions and comprehension questions again

☐ I will continue with my investment decision

[If they choose the ‘take me back option’, they go back to the “Welcome to our Study” screen. Everything is the same, except when they get to this page again, which would count as their second attention check failed and they would not be able to continue with the study.]

[passed attention check] passed it.

Thank you for paying attention while reading the instructions and being fully focused on this study.

The next screen will ask you about your investment decision.

[everyone] Remember that your answers are very important to us, and part of your payoff in this study depends on them!

[Attention check feedback 2. Participants get this screen instead of 'Attention screen feedback 1" if they chose "take me back to see the instructions…" on that page. If they choose to go back, the failed attention check count restarts.]

Attention check

[10 min timer]

The previous question was an attention check and you

[if they passed the attention check]

passed it.

[if they did not pass the attention check]

did not pass it.

[everyone]

Thank you for paying attention while reading the instructions and being fully focused on this study.

The next screen will ask you about your investment decision.

Remember that your answers are very important to us, and part of your payoff in this study depends on them!

Investment Decision

[10 min timer]

[The following box is only shown if they click on 'Show payment info reminder' button]

Remember that,

[T1] If your group is rewarded with **[T1: order1]** 03 **/ [T1: order2]** 0.5 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be

[T1: order1] 0.3 / [B_0.4] 0.4 / [B_0.5, T1: order2 and T2, T2x, T3] 0.5 x (total Taler invested in the common project by everyone)

- Each participant's total earnings will be
  (10 – own investment) + [T1: order1] 0.3 / [B_0.4] 0.4 / [B_0.5, T1: order2 and T2, T2x, T3] 0.5 x (total group contribution to the common project)

[T1] If your group is rewarded with **[T1: order1]** 0.5 **/ [T1: order2]** 0.3 Taler for each Taler contributed to the common project:

- Each participant's income from the common project will be
  [T1: order1] 0.5 / [T1: order2] 0.3 x (total group contribution to the common project)
- Each participant's total earnings will be
  (10 – own investment) + [T1: order1] 0.5 / [T1: order2] 0.3 x (total group contribution to the common project)

- **[T2]** There is a 20% chance that the investment to the common project of each of your group members (including yourself) is lost.

There are 10 Taler in your private account.

From those, how many would you like to invest in the common project?

[dropdown list 0-10]

Contributions to the Common Project

You are player **[A, B, C, D – assigned as they 'join the group']**.	**[this column shown in T2x only; if contributions were not lost, then "—", if they were, then the amount is displayed]**
Common Project Contribution	**Lost Contributions**
Player A: **[A's contribution]** Taler	--
Player B: **[B's contribution]** Taler	7 Taler
Player C: **[C's contribution]** Taler	--
Player D: **[D's contribution]** Taler	--

You and your group members will receive **[theta]** Taler for each Taler in the common project.

Each group member's share of the income from the common project: **[theta]** x **[group.total_contribution]** = **[group.individual_share]** Taler.

Some questions about you before you finish

What is your year of birth?

[drop down list: 1920 to 2004 including “I prefer not to answer”]

What is your gender?

- Female
- Male
- Other (describe in the box below if you wish)
- I prefer not to answer

How well would you say you are managing financially these days?

- Living comfortably
- Doing alright
- Just about getting by
- Finding it quite difficult
- Finding it very difficult
- I do not know
- I prefer not to answer

Have you ever taken a course or learnt about economics and/or game theory? Yes/No

[Attention check]

Please, choose the answer ‘a little’ below.

- A great deal
- A lot
- A moderate amount
- A little
- None at all

[Trust in institutions]

- 1 to 10 scale with 1 - No confidence at all, 10 - Full confidence, and numbers 2 to 9 without labels in all questions.

Do you have confidence in the judicial system and courts or not?

Do you have confidence in the national government or not?

Do you have confidence in the honesty of elections or not?

[Political views]

In political matters, people talk of “the left” and “the right”. Generally speaking, how would you place your views on this scale?

- 10-point horizontal scale with “Left” and “Right” labels

Earnings

[5 min timer]

Remember that your total earnings from this study are the sum of:

- The Taler left in your private account
- Your share of the income from the common project, which is computed as **[theta] [this value changes in each condition. B0.5 = 0.5. T1 = (0.5, 0.3; 0.5, 0.5], T2, T2x = 0.5]** *x (total group contribution to the common project)*
- Your participation fee

Your total earnings from this study are *10 – **[investment]** + **[individual_share]*** = **[payoff_withoutshowupfee]** Taler, which are equivalent to £**[payoff_usd]** + your participation fee (£1). That is, a total of **[payoff]**.

[More questions]

[5 min timer]

The following questions will have no effect on you final payment. We will not reject your submission on the basis of these questions either. Please answer truthfully.

1. Did you trust the instructions for the tasks you have gone through to be accurate and true while reading them? **[Yes/No]**
2. Did you find that you could understand fully the instructions with no misunderstanding? **[Yes/No]**
3. Did you complete all the tasks on your own, or did you get assistance from someone else? **[On my own/With someone's help]**
4. Were you focusing only at the tasks while completing them, or were you also doing something else on your browser or device? **[Tasks only/Something else]**

Study Completed

Thank you for taking part in this study!
Your completion code is **[XXXXXXXXXX]**.
Click here to be redirected to Prolific's completion page.

PART 2: MPL & SVO

[Introduction]

Welcome to this study!

You will complete two tasks.

At the end of the study, one task will be randomly selected and your bonus payment will depend on the choices you or someone else made in a question within that task, and sometimes luck.

Please, pay attention while completing this study. Your answers are important for our research.

____

Task 1

[The order of tasks 1 and 2 is randomised]

In this task you have to make 10 choices between "Option A" and "Option B", one per row. The payoffs of the two options are the same in all choices, but the chances of getting the high payoff in each option will vary in each choice.

If this task is randomly chosen for payment, one of the choices will be randomly selected. Then, we will draw a virtual 10-sided dice to determine whether you get the high or the low payoff in that option as your bonus payment.

	Option A	Option B	
£2.00 with a probability of 1/10, £1.60 otherwise	○	○	£3.85 with a probability of 1/10, £0.10 otherwise
£2.00 with a probability of 2/10, £1.60 otherwise	○	○	£3.85 with a probability of 2/10, £0.10 otherwise
£2.00 with a probability of 3/10, £1.60 otherwise	○	○	£3.85 with a probability of 3/10, £0.10 otherwise
£2.00 with a probability of 4/10, £1.60 otherwise	○	○	£3.85 with a probability of 4/10, £0.10 otherwise
£2.00 with a probability of 5/10, £1.60 otherwise	○	○	£3.85 with a probability of 5/10, £0.10 otherwise
£2.00 with a probability of 6/10, £1.60 otherwise	○	○	£3.85 with a probability of 6/10, £0.10 otherwise
£2.00 with a probability of 7/10, £1.60 otherwise	○	○	£3.85 with a probability of 7/10, £0.10 otherwise
£2.00 with a probability of 8/10, £1.60 otherwise	○	○	£3.85 with a probability of 8/10, £0.10 otherwise
£2.00 with a probability of 9/10, £1.60 otherwise	○	○	£3.85 with a probability of 9/10, £0.10 otherwise
£2.00 with a probability of 10/10, £1.60 otherwise	○	○	£3.85 with a probability of 10/10, £0.10 otherwise

____

Task 2

In this task, you will make 6 decisions about allocating points between you and another other person. Every point is worth £0.01.

Each option will be presented as follows. You will have to choose the combination of points for yourself and points for the other person that you prefer in each question.

You Receive [Points for you here]
|
|
Other Receives [Points for the other person here]

If this task is selected for payment, you will be randomly paired with another person. One of you will be randomly allocated the role of decider, and the other one the receiver.
Neither you nor the other person will ever know each other's identity.

One of the six questions will be chosen at random.
If you are the decider, you and the other person will be paid according to your choice in that question. If you are the receiver, you and the other person will be paid according to the other person's choice in that question.

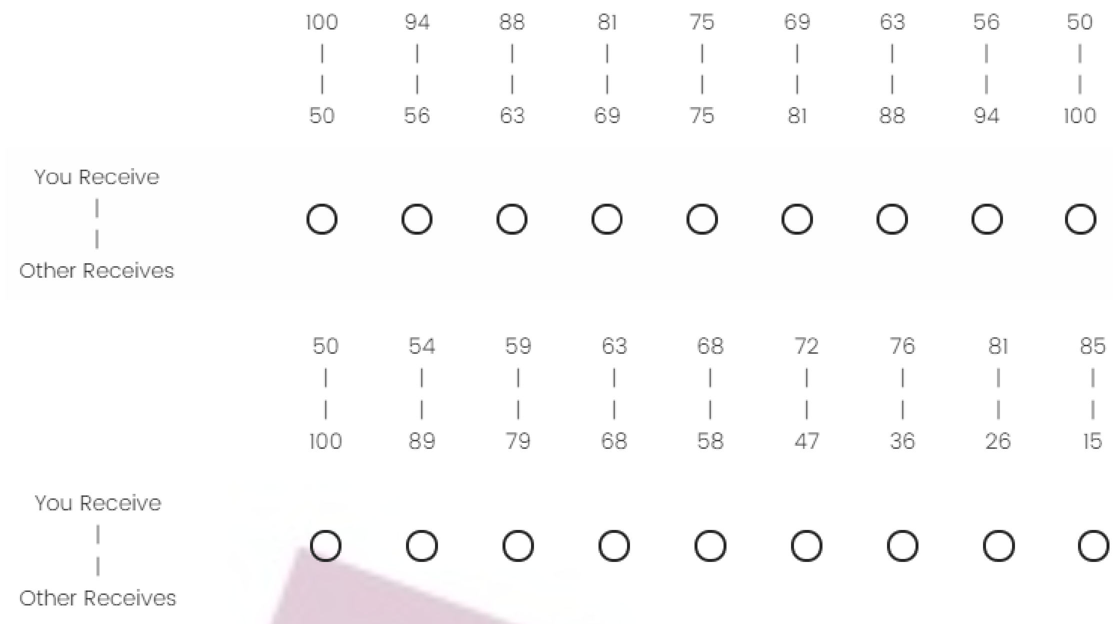
[The order of the following questions is randomised at the participant level.]

You Receive	50	54	59	63	68	72	76	81	85
Other Receives	100	98	96	94	93	91	89	87	85
	○	○	○	○	○	○	○	○	○

You Receive	85	85	85	85	85	85	85	85	85
Other Receives	85	76	68	59	50	41	33	24	15
	○	○	○	○	○	○	○	○	○

You Receive	85	87	89	91	93	94	96	98	100
Other Receives	15	19	24	28	33	37	41	46	50
	○	○	○	○	○	○	○	○	○

You Receive	100	98	96	94	93	91	89	87	85
Other Receives	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	◉



____

[Payment]

Your bonus payment will depend on **[task chosen at random]**. If you would like more details about the question that was randomly chosen for payment within that task, get in touch after you have received your bonus payment.

If you have any comments about this study, write them in the box below.
[textbox here]

Comprehension Questions by treatment

#	Variable	Question	Base line	Risk T1	PD T2	NoPD T2x	PD-Part T3
1	attention_t1	[Attention check] What is the chance that you and the other members of your group receive 0.3 Taler for every Taler invested into the common project?		x			
2	q3_words	How many Taler would each group member receive as income from the common project? (text answers)	x	x	x	x	x
3	q4_words	Your total earnings would amount to… (text answers)	x	x	x	x	x
4	quiz3	How many Taler will each group member receive as income from the common project? (numerical answers)	x	x	x	x	x
5	quiz4	Your total earnings would amount to… (numerical answers)	x	x	x	x	x
6	attention_taler	[Attention check] Choose five taler in this question…	x	x	x	x	x
7	quiz_chance_other	What is the chance that … the investment of each of your group members is not lost?			x	x	x
8	quiz_chance_you	[Attention check] What is the chance that … your investment is lost?			x	x	x
9	quiz_investment	If you invested 7 Taler but … got lost, what contribution would appear from you as Player B on the Common Project Contributions table?			x	x	x
10	attention_check	[Attention check] This is a screen to ensure you are attentive to the instructions and the questions asked during this task. … What is the answer to the question above?	x	x	x	x	x
11	attention_check2	[Attention check] Please, choose the answer 'a little' below.	x	x	x	x	x

NB: "_2" is added to the question label when participants encounter it if they click to "go back and revisit instructions"

B Expected Marginal per Capita Returns Calculation

In what follows, we assume a public good game of N players. Each player, i, decides their level of contribution, c_i, out of their respective endowment which is equal to E. The marginal per capita return (MPCR) of the public good is given by parameter θ. Thus, the payoff for each player is given by:

$$\pi_i(c_i) = E - c_i + \theta \sum_{j=1}^{N} c_j$$

B.1 Whole contribution lost

Assume now that a player i's *intended* contribution, c_i, towards the public good may be lost with probability γ. This implies that the expected contribution for each player i is:

$$E(c_i) = \gamma * 0 + (1-\gamma) * c_i = (1-\gamma)\, c_i$$

In expectation, the common pool is equal to summing up all expected contributions. For N players:

$$\sum_{i=1}^{N} (1-\gamma)\, c_i$$

Hence, the expected utility for each player i is given by:

$$EU_i(c_i) = E - c_i + \theta \sum_{i=1}^{N} (1-\gamma)\, c_i \implies EU_i(c_i) = E - c_i + \theta\,(1-\gamma) \sum_{i=1}^{N} c_i$$

That is, the expected MPCR of the public good when accounting for the probability of contributions getting lost is:

$$\theta\,(1-\gamma)$$

With $\theta = 0.5$ and $\gamma = 0.2$ (experimental parameters used in our *PD* treatments) we obtain expected θ:

$$0.5\,(1-0.2) = 0.4$$

B.2 Contribution partially lost

Assume now a situation where a player i's share α of their *intended* contribution, c_i, towards the public good may be lost. Once again, this happens with probability γ. This implies that the expected contribution for each player i is:

$$E(c_i) = \gamma * (1-\alpha)\, c_i + (1-\gamma) * c_i = (1-\alpha\gamma)\, c_i$$

In expectation the common pool is equal to summing up all expected contributions. For N players:

$$\Sigma_{i=1}^{N} (1-\alpha\gamma)\, c_i$$

Hence, expected utility for each player i:

$$EU_i(c_i)i = E - c_i + \theta\Sigma_{i=1}^{N}\left(1-\alpha\gamma\right)c_i \implies EU_i(c_i) = E - c_i + \theta\left(1-\alpha\gamma\right)\Sigma_{i=1}^{N}c_i$$

That is, the expected MPCR of the public good when accounting for partial contributions getting lost:

$$\theta\left(1-\alpha\gamma\right)$$

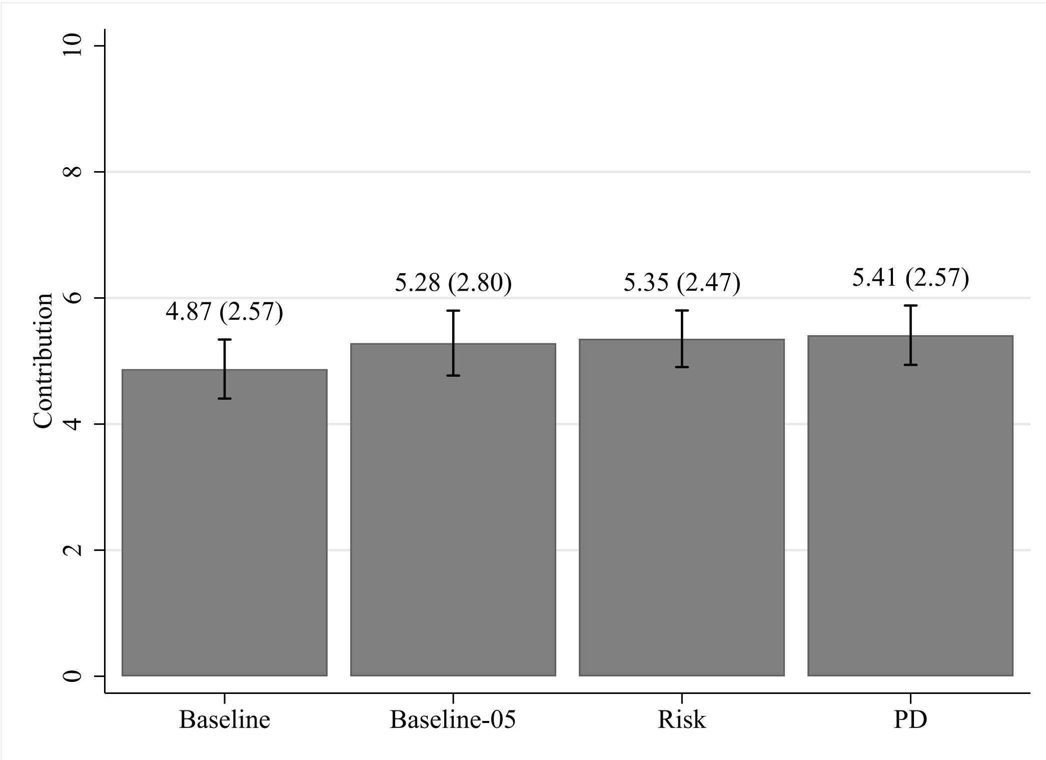
With the experimental parameters used in our *PD-Partial* treatment, $\theta = 0.5$, $\gamma = 0.3$, and $\alpha = \frac{2}{3}$, we obtain expected θ:

$$0.5\left(1-\frac{2}{3}*0.3\right) = 0.4$$

C Additional Analysis

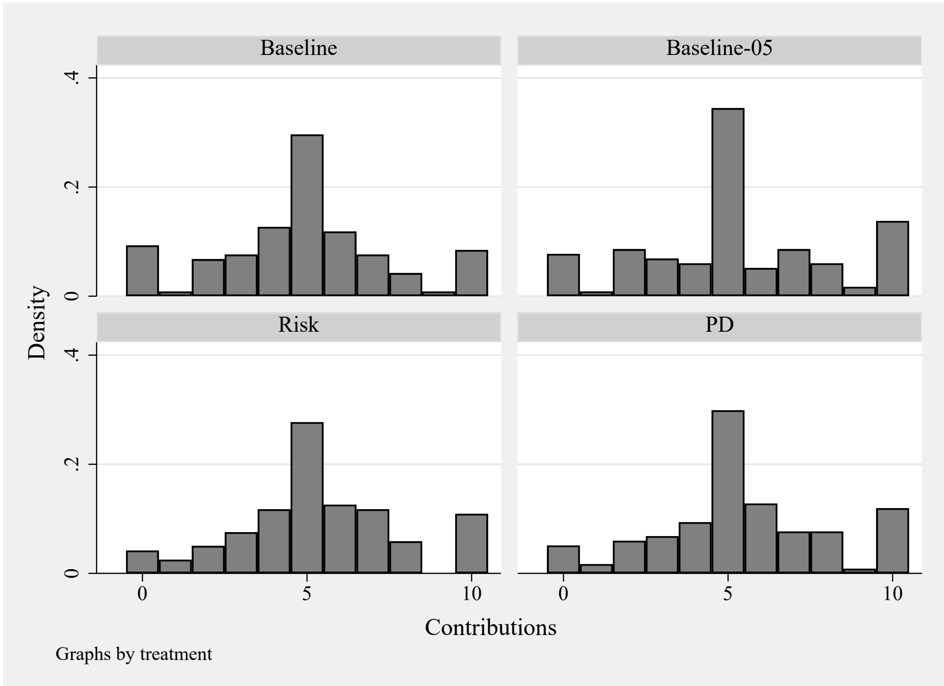
C.1 Study 1

Figure A.1: Average Contribution Across Treatments (incl. *Baseline-05*).



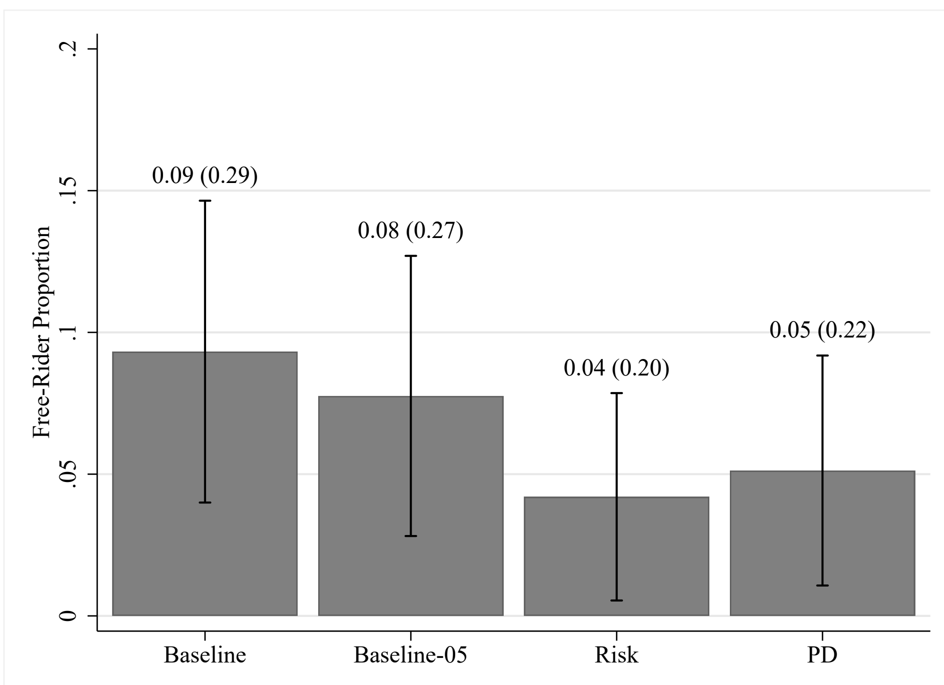
Note: We denote the average *intended* contribution above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

Figure A.2: Distribution of Contributions by Treatment



Note: We report the distribution of *intended* contributions. PD = Plausible Deniability. Non-parametric tests conclude that there are no significant differences in pairwise comparisons of each treatment distribution (all $p > 0.11$, exact MWU test).

Figure A.3: Proportion of Free-riders Across Treatments (incl. *Baseline-05*).



Note: We denote the proportion above each bar, with the standard deviation in brackets. Bands represent the 95% confidence interval. PD = Plausible Deniability.

Table A.1: Regression Analysis of Treatment Effects on Contributions & Free-Riding.

	Amount Sent (OLS)		Free-Riding (Logit)	
	b/se	b/se	OR/se	OR/se
ref cat	*Baseline*	*Risk*	*Baseline*	*Risk*
Risk	0.2307		0.4175	
	(0.3906)		(0.3069)	
Plausible Deniability	0.3918	0.1860	0.5617	0.8625
	(0.4439)	(0.3842)	(0.4481)	(0.7094)
Controls	Yes	Yes	Yes	Yes
N	288	194	288	194

Note: Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right); No. correct answers in comprehension quiz. * p<0.10, ** p<0.05, *** p<0.01

Table A.2: Regression Analysis of Treatment Effects on Amount Sent & Free-Riding (First Robustness Check).

	Amount Sent (OLS)		Free-Riding (Logit)	
	b/se	b/se	OR/se	OR/se
ref cat	*Baseline*	*Risk*	*Baseline*	*Risk*
Risk	0.2148		–0.6449	
	(0.4904)		(0.9515)	
Plausible Deniability	0.2610	0.0512	0.3383	2.9588
	(0.5112)	(0.5016)	(0.9000)	(2.9882)
Controls	Yes	Yes	Yes	Yes
N	158	111	158	111

Note: Excluding participants that got less than the median share of correct answers in the comprehension quiz questions at first try. The median for the correct answer share is 100% for the *Baseline* and *PD* treatments, while it is 80% for the *Risk* treatment. Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right). * p<0.10, ** p<0.05, *** p<0.01

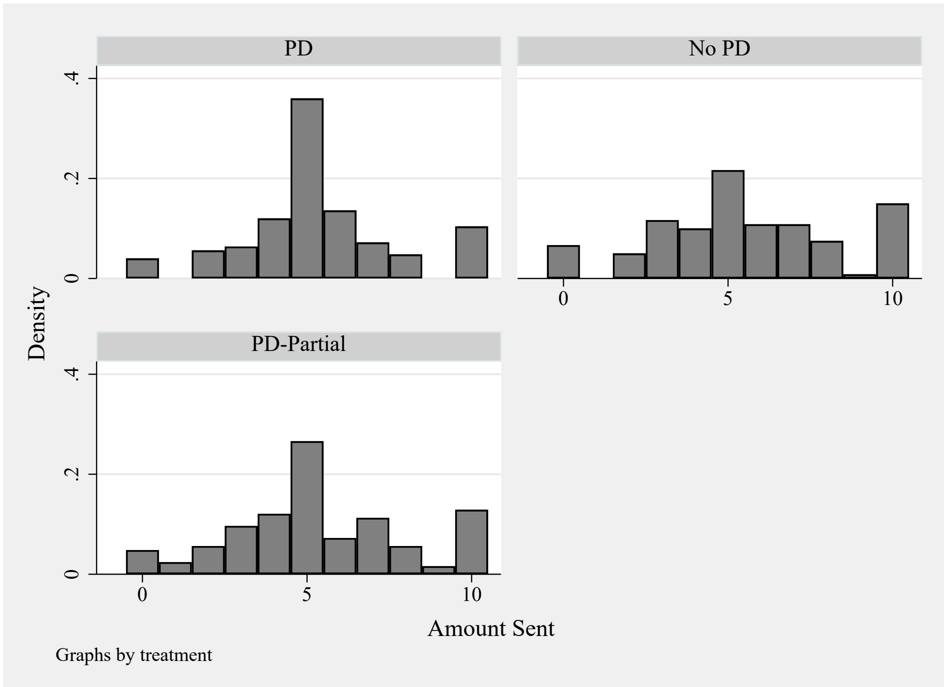
Table A.3: Regression Analysis of Treatment Effects on Amount Sent & Free-Riding (Second Robustness Check).

	Amount Sent (OLS)		Free-Riding (Logit)	
	b/se	b/se	OR/se	OR/se
ref cat	*Baseline*	*Risk*	*Baseline*	*Risk*
Risk	0.4053		–0.2362	
	(0.5492)		(0.9722)	
Plausible Deniability	0.3099	–0.1686	0.3794	2.1307
	(0.5322)	(0.5558)	(0.9104)	(2.1325)
Controls	Yes	Yes	Yes	Yes
N	139	93	139	93

Note: Excluding non-attentive participants, participants that did not get all the comprehension quiz questions correct at first try, and participants that self-reported not to fully understand the instructions. Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right). * $p<0.10$, ** $p<0.05$, *** $p<0.01$

C.2 Study 2

Figure A.4: Distribution of Contributions by Treatment



Note: We report the distribution of *intended* contributions. PD = Plausible Deniability. Non-parametric tests conclude that there are no significant differences in pairwise comparisons of each treatment distribution (all $p > 0.56$, exact MWU test)

Table A.4: Regression Analysis of Treatment Effects on Contributions & Free-Riding.

	Amount Sent (OLS)			Free-Riding (Logit)		
	b/se	b/se	b/se	OR/se	OR/se	OR/se
ref cat	*PD*	*PD*	*NoPD*	*PD*	*PD*	*NoPD*
No Plausible Deniability	–0.1261			3.2357		
	(0.3797)			(3.2057)		
Partial Pl. Deniability		–0.2609	–0.0563		8.9643	0.4419
		(0.3809)	(0.4441)		(16.5386)	(0.4496)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	159	154	149	83	76	75

Note: Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right); No. correct answers in comprehension quiz. * p<0.10, ** p<0.05, *** p<0.01

Table A.5: Regression Analysis of Treatment Effects on Amount Sent & Free-Riding (First Robustness Check).

	Amount Sent (OLS)			Free-Riding (Logit)		
	b/se	b/se	b/se	OR/se	OR/se	OR/se
ref cat	*PD*	*PD*	*NoPD*	*PD*	*PD*	*NoPD*
No Plausible Deniability	–0.2542			21.6363		
	(0.5092)			(59.6291)		
Partial Pl. Deniability		–0.3432	0.1384		5.9678	0.8962
		(0.5375)	(0.6113)		(11.4860)	(1.5985)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	95	94	81	45	41	42

Note: Excluding participants that got less than the median share of correct answers in the comprehension quiz questions at first try. The median for the correct answer share is 100% for the *PD-Partial* and *NoPD* treatments, while it is 75% for the *PD* treatment. Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right). * p<0.10, ** p<0.05, *** p<0.01

Table A.6: Regression Analysis of Treatment Effects on Amount Sent & Free-Riding (Second Robustness Check).

	Amount Sent (OLS)			Free-Riding (Logit)		
	b/se	b/se	b/se	OR/se	OR/se	OR/se
ref cat	*PD*	*PD*	*NoPD*	*PD*	*PD*	*NoPD*
main						
No Plausible Deniability	–0.2200			21.6363		
	(0.5418)			(59.6291)		
Partial Pl. Deniability		–0.2204	0.1929		20.8951	1.1962
		(0.5870)	(0.6763)		(51.4066)	(2.2662)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	82	77	71	45	37	38

Note: Excluding non-attentive participants, participants that did not get all the comprehension quiz questions correct at first try, and participants that self-reported not to fully understand the instructions. Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right). * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A.7: Regression Analysis of Treatment Effects on Amount Sent & Free-Riding (Third Robustness Check).

	Amount Sent (OLS)			Free-Riding (Logit)		
	b/se	b/se	b/se	OR/se	OR/se	OR/se
ref cat	*PD*	*PD*	*NoPD*	*PD*	*PD*	*NoPD*
No Plausible Deniability	0.1433			2.3729		
	(0.4383)			(2.9013)		
Partial Pl. Deniability		0.1345	0.0461		7.0863	0.7713
		(0.4364)	(0.5229)		(12.4190)	(1.3379)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	113	119	102	59	57	36

Note: all participant that did not answer correctly the question that asked about the number that would appear on the contributions table if they were unlucky and their contribution was lost (28% of the sample). Controls: Female Dummy; Age; No. safe choices (MPL); SVO; Trust questions (1-10 from no confidence at all to full confidence): trust in judicial system, trust in government, trust in elections; Political views (1-10 from left to right). * $p<0.10$, ** $p<0.05$, *** $p<0.01$